The Warwick Holistic Health Questionnaire.

The Development and Validation of a

Patient-Reported Outcome Measure for Craniosacral Therapy: a Mixed Methods Study.

By

Nicola Brough

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor in Philosophy in Health Sciences

University of Warwick, Warwick Medical School

September 2017

Table of Contents

List of figures **11**

List of tables **12**

List of equations **14**

Abbreviations **15**

Declaration **17**

Research training **18**

Abstract **21**

Acknowledgements **22**

Dedication **23**

1. Introduction **24**

1.1 Chapter Overview 24
1.1.1 Aims and objectives 24
1.1.2 Background 25
1.1.3 Defining CAM 25
1.1.4 Concepts of health: a complementary and integrative medicine perspective 26
1.1.5 Craniosacral Therapy (CST) 27
1.1.6 Concepts of health from the bio-medicine perspective 31
1.1.7 Inter-disciplinary and historical context 32

1.2 Outcome measurement in health 33
1.2.1 Conceptual frameworks: reflective and formative models 36
1.2.2 Conceptual framework for CAM 36
1.2.3 IN-CAM outcomes database 37
1.2.4 Health related quality of life 38
1.2.5 Spiritual wellbeing 38
1.2.6 The relationship between HRQoL and wellbeing 39
1.2.7 Measuring outcomes for CST 40

1.3 Typology of Measures 41

1.3.1 Generic measures 41
1.3.2 Utility measures 41
1.3.3 Specific measures 42
1.3.4 Individualised measures 42
1.3.5 Patient reported outcomes 42
1.3.6 Chapter summary 43

2 Literature Review 44

2.1 Introduction to literature review 44

2.2 Inclusion criteria for a CST PRO 45

2.3 Systematic search to identify PROs for CST 45
2.3.1 Search 1: identifying PROs for CST 45
2.3.2 Results of search 1 47

2.4 Search 2: identifying CST studies using PROs for evaluation 47
2.4.1 Methods for search 2 47
2.4.2 Results of search 2 49
2.4.3 Search 2 summary 58

2.5 Search 3: identifying and appraising PROs developed for CAM 58
2.5.1 Methods for search 3 58
2.5.2 Results of search 3 60
2.5.3 A) PROs developed for CAM from the Hunter review 63
2.5.4 Measure Your Medical Outcome Profile (MYMOP) 65
2.5.5 B) Other PROs developed for CAM identified from systematic search 66
2.5.6 Self-Assessment of Change (SAC) 68

2.6 Consultation with experts 69
2.6.1 Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) 69

2.7 A synopsis of the eight questionnaires reviewed in this thesis 70

2.8 Methodological issues of searching for PROs for CST and CAM 74

2.9 Chapter summary 74

3 Methodology 76

3.1 Chapter Overview 76
3.1.1 Ethical issues 76
3.1.2 Bias 76
3.1.3 Reflexivity 77
3.1.4 Sampling 77
3.1.5 PRO development 77
3.1.6 Reporting standards and guidelines 78

3.2 Mixed methods 78

3.3 Theoretical perspectives 81

3.4 Qualitative methods rationale 82
3.4.1 Focus groups 82
3.4.2 Interviews 83
3.4.3 Semi-structured interviews 84
3.4.4 Cognitive interviews 84

3.5 Quantitative methods rationale 85
3.5.1 Classical Test Theory 86
3.5.2 Item Response Theory (IRT) 87
3.5.3 Rasch models 87
3.5.4 Limitations of IRT/Rasch 88

3.6 Choice of model 88

3.7 Reliability 89
3.7.1 Internal reliability 89
3.7.2 Repeatability reliability 89
3.7.3 Intraclass correlation coefficients (ICC) 90
3.7.4 Measurement error 91

3.8 Validating a PRO 91
3.8.1 Content and face validity 91
3.8.2 Criterion validity 91
3.8.3 Construct validity 92
3.8.4 Structural validity 92
3.8.5 Convergent validity 92
3.8.6 Cross cultural validity 92

3.9 Responsiveness 93

3.10 Interpretability 93
3.10.1 Distribution of scores of the instrument 93
3.10.2 Evaluating edge effects 94
3.10.3 Interpreting change 94
3.10.4 Smallest detectable change 95
3.10.5 Minimal important change 95
3.10.6 Other considerations regarding interpretability 95

3.11 Chapter Summary 96

4 Methods 97

4.1 Ethics and governance 97
4.1.1 Ethical approval 97
4.1.2 Rigour and validity 97
4.1.3 Bias 97

4.2 Study design for PRO Development 98

4.3 Establishing a conceptual framework for CST 99

4.4 The importance of having a conceptual framework 99

4.5 Concepts of CST 99

4.6 Conceptual framework of CST outcomes and draft questionnaire assessment 100
4.6.1 Sampling strategy 100
4.6.2 Recruitment 100
4.6.3 Consultation 101
4.6.4 Inclusion and Exclusion criteria 101

4.7 Data collection 102

4.8 Practitioner focus groups 102
4.8.1 Researcher's checklist 103
4.8.2 Facilitation of the focus groups 104
4.8.3 Schedule for focus groups: 105
4.8.4 Debriefing notes 106
4.8.5 CST users' focus groups - evaluating a working conceptual framework of CST outcomes 106

4.9 Draft PRO questionnaire 106
4.9.1 Determined population 107
4.9.2 Determined use 107
4.9.3 Naming the questionnaire 107
4.9.4 Construction of a draft version of the questionnaire 107
4.9.5 Evaluating the draft PRO 109
4.9.6 Data analysis for semi-structured interviews 110
4.9.7 Consensus meeting with CST practitioners 110
4.9.8 Pre-testing the draft PRO (52-item WHHQ) 111
4.9.9 Data storage 112

4.10 Testing the draft questionnaire in a group of CST users 112
4.10.1 Assessing the measurement properties 112
4.10.2 Recruitment 112
4.10.3 Inclusion and exclusion criteria of CSTA participants 113
4.10.4 Data collection and method of administration 113
4.10.5 Data handling and cleaning 113
4.10.6 Scoring rules and missing data for WHHQ and comparator measures 113
4.10.7 Floor and ceiling effects 115

4.11 Measurement model 115
4.11.1 Descriptive statistics 115
4.11.2 Psychometrics 115
4.11.3 Intra class correlation coefficients 116

4.12 Measurement error 117

4.13 Structural validity 117
4.13.1 Exploratory factor analysis 117
4.13.2 Model refinement 119
4.13.3 Assessing content and face validity 119

4.14 Evaluating the measurement properties of the final version of the WHHQ 120
4.14.1 Construct validity 120
4.14.2 Hypothesis testing 120
4.14.3 Strength or magnitude of the relationship 120
4.14.4 Bland and Altman plots 121

4.15 Assessing the 25-item WHHQ's ability to detect change 121
4.15.1 Responsiveness 121

4.16 The smallest detectable change (measurement error) 121

4.17 Minimal important change (MIC) 122

4.18 Chapter summary 122

5 Results: Evaluating the conceptual framework of CST outcomes 124

5.1 Introduction 124

5.2 Focus groups with CST practitioners 126
5.2.1 Demographics of focus groups 1 and 2 127

5.3 Focus group 1 127
5.3.1 Purpose and function of the conceptual framework 127
5.3.2 Layout and design 128
5.3.3 Content: domain, sub-domain level and component level 128
5.3.4 Summary of focus group 1 130

5.4 Focus group 2 130
5.4.1 Purpose and function of the conceptual framework 130
5.4.2 Layout and design 130
5.4.3 Content: domains, sub-domains, items 131
5.4.4 Summary of focus group 2 132

5.5 Revisions to the conceptual framework after focus groups 1 and 2 132

5.6 Focus group 3 – CST users 133
5.6.1 Demographics of focus group 3 133
5.6.2 Layout and design of conceptual framework 134
5.6.3 Content of conceptual framework: domain, sub-domain level and item level 134
5.6.4 Summary of focus group 3 136

5.7 Revisions to conceptual framework based on CST users' perspectives 136

5.8 Chapter summary 140

6 Pre-testing preliminary WHHQ 141

6.1.1 Demographics of the sample – semi structured interviews, rounds 1 and 2 141

6.2 Consensus meeting with CST practitioners 141

6.3 Design, layout and instructions 141

6.4 Recall period – 2 weeks 142

6.5 Name of the questionnaire 142

6.6 Response options 143

6.7 Content 144

6.8 Pre-testing of 52-item WHHQ - cognitive interviews, round 3 151

6.9 Demographics of the sample, round 3 151

6.10 Interviews 151

6.11 Design, layout, instructions 151

6.12 Response options, round 3 152

6.13 Content, round 3 152

6.14 Chapter summary 152

7 Psychometric testing of the 52-item WHHQ 154

7.1 CST practitioners who collected data 154

7.2 Demographics of sample 1 154

7.3 Gender 154

7.4 Age of participants 155

7.5 Number of CST sessions undertaken and year of first CST session 155

7.6 Self-reported overall wellbeing 157

7.7 Reasons for having sessions 157

7.8 Construct validity 158

7.9 Exploratory factor analysis results 160

7.10 Item reduction using exploratory factor analysis 160

7.11 Final reduced item Model 163

7.12 Face and content validity cross check 165

8 Structural validity 166

8.1 Pattern and correlation matrices 166

8.2 25-item WHHQ 169

9 Psychometric evaluation of the 25-item WHHQ 170

9.1 Patient demographics of sample 2 171

9.2 Number of CST sessions undertaken and year of first CST session 171

9.3 Reasons for having CST sessions 172

9.4 Timing of CST sessions 173

9.5 Self-report of health status 173

9.6 Psychometric properties of 25-item WHHQ 175

9.7 Reliability 175
9.7.1 Distribution of the scale 177

9.8 Test of repeatability 177

9.9 Responsiveness 179

9.10 The effect size (ES) 180

9.11 Standardised response mean (SRM) 181

9.12 Interpretability 181
9.12.1 Minimal important change (MIC) 181

9.13 Convergent validity 184

9.14 Chapter summary 187

9.15 Summary of results 188
9.15.1 Conceptual framework of CST outcomes 189
9.15.2 Changes to the conceptual framework made through consensus 189

9.16 The literature review of PROs developed for CST and CAM 192

9.16.1 Results: qualitative 196
9.16.2 Results: quantitative 197
9.16.3 Evaluating the measurement properties of the 25-item WHHQ 198
9.16.4 Acceptability 198
9.16.5 Reliability 198
9.16.6 Validity 199
9.16.7 Responsiveness 199

10 Discussion 200

10.1 Chapter Overview 200

10.2 Summary of the findings 200

10.3 Adequacy of research methods and implications for instrument validity 201
10.3.1 Literature review 201
10.3.2 Development and validation of the conceptual framework 202
10.3.3 Development and validation of the WHHQ 203
10.3.4 Determining the format of items and response options 205
10.3.5 Efforts to limit response bias 206
10.3.6 Measurement theory applied in this study 206
10.3.7 Determining the scoring 207
10.3.8 Semi-structured interviews 207
10.3.9 Data collection (qualitative) 207
10.3.10 Organising and analysing data 207
10.3.11 Reflexivity 208

10.4 Findings in relation to the literature review 210
10.4.1 Health related quality of life 210
10.4.2 Wellbeing 211
10.4.3 Taking responsibility for oneself 213
10.4.4 Social aspects of health 213
10.4.5 MYCaW coding guidelines 215

10.5 The contribution this work offers 216

10.6 Strengths of this work 222

10.7 Limitations of this work 223
10.7.1 Ethical Tensions of being a Practitioner/Researcher 224

10.8 Questionnaire implementation 226
10.8.1 Managing the implementation of the WHHQ for the CSTA 227

10.9 Future research 227
10.9.1 Refining the 25-item WHHQ 228
10.9.2 Research setting 228
10.9.3 Evaluating clinical practice 228
10.9.4 Electronic version of the 25-item WHHQ 229
10.9.5 Use in other populations 229

11 Conclusions **230**

12 Appendices **232**

12.1 Appendix 1: Ethics approval letter 232

12.2 Appendix 2: Recruitment poster 233

12.3 Appendix 3: Participant information leaflet 234

12.4 Appendix 4: Revised working conceptual framework of CST outcomes 239

12.5 Appendix 5: List of draft items (73-WHHQ) 241

12.6 Appendix 6: Interview schedule for cognitive interviews 245

12.7 Appendix 7: A2 Conceptual Framework of CST outcomes 247

Reference List **248**

List of figures

Figure 1: CST session vault hold (image used with permission) 28
Figure 2: Wilson and Cleary conceptual model for health-related quality of life 35
Figure 3: Framework of health outcome domains relevant to CAM research and practice 37
Figure 4: Results of the systematic search for PROs used in studies evaluating the effects of CST 50
Figure 5: Flow of information through the systematic search of PROs for CAM 60
Figure 6: Evolution of instrument development (adapted from FDA Guidance) 80
Figure 7: Working draft of Conceptual Framework of CST outcomes 125
Figure 8: Revised version of the conceptual framework of CST outcomes 138
Figure 9: Number of practitioners and data collection locations 154
Figure 10: Sample 1, year of first CST session 156
Figure 11: Scree plot for 19 items (Model 33) 163
Figure 12: Location of UK and non-UK practitioners for sample 2 170
Figure 13: Sample 2, year of first CST session 172
Figure 14: Days between baseline and T2 173
Figure 15: Histogram of baseline data 177
Figure 16: Plot showing 25-item WHHQ and SF-12v2 PCS scores 185
Figure 17: Plot showing 25-item WHHQ and SF-12v2 MCS scores 186
Figure 18: Phases in the development & psychometric evaluation of the WHHQ 188

List of tables

Table 1: Domains and outcomes of importance for CST 30
Table 2: Criteria for selecting questionnaires 45
Table 3: Search 1 search terms 46
Table 4: Search 1 search strategies and results – Medline 46
Table 5: Search 2 identifying PROs used in CST studies 47
Table 6 Search results identifying CST studies using PROs for evaluation 48
Table 7: CST studies and outcome measures used 52
Table 8: SF-36 internal consistency and test-retest reliability 56
Table 9: Search 3, search terms, strategies and results (2016 search) 59
Table 10: Internet search terms 62
Table 11: Candidate PROs (non-CAM) 72
Table 12: Candidate PROs developed for CAM assessed in this thesis 73
Table 13: Key differences between think-aloud and probing techniques 85
Table 14: The assumptions underpinning classical test theory 86
Table 15: Conceptual Framework of CST Outcomes Domain Definitions 126
Table 16: Revisions made to working conceptual framework after focus groups 1 and 2 133
Table 17: Revisions to conceptual framework of CST outcomes after focus group with CST users 137
Table 18: Working conceptual framework of CST outcomes after revisions 139
Table 19: Summary of item changes per interview rounds based on patient reports and consensus meeting with CST practitioners 145
Table 20: Age of participants by gender 155
Table 21: Sample 1, number of CST previous sessions received by gender 156
Table 22: Sample 1, self-reported overall wellbeing by gender 157
Table 23: Sample 1 - Reasons for having CST sessions 158
Table 24: Reasons for attending CST sessions by gender 158
Table 25: 52-Item WHHQ hypothesized domains and representing statements 159
Table 26: Item reduction list 161
Table 27: Total variance explained 163
Table 28: Pattern Matrix of Model 34 – draft WHHQ 2: 19 items of 4 factors 164
Table 29: Items to be reconsidered for inclusion in the WHHQ 165
Table 30: Draft 3 WHHQ: Pattern Matrix 25-item WHHQ 167
Table 31: Component Correlation Matrix 25-item WHHQ 168
Table 32: 25-Item WHHQ with response options 169
Table 33: Sample 2, age by gender 171
Table 34: Sample 2 number of CST sessions undertaken by gender 171
Table 35: Sample 2 (baseline) reasons for having CST 173
Table 36: Sample 2, self-rated overall wellbeing by gender at baseline 174
Table 37: Sample 2, self-rated overall wellbeing by gender at T2 174
Table 38: Cronbach's alpha with item removed 176
Table 39: CST users whose health remained stable in-between baseline and T2 178

Table 40: Paired sample T-test WHHQ baseline – T2 for those who reported no change 178
Table 41: Intra-class correlation calculated model: absolute agreement 179
Table 42: Descriptive statistics 25-item WHHQ at baseline and T2 for participants who changed 179
Table 43: Change scores at baseline and T2 per sub-sample. 180
Table 44: Mean change score of four PROs per the health status anchor 182
Table 45: PRO characteristics and scores at baseline and T2 N=105 182
Table 46: PRO characteristics of scores at baseline and T2 n=105 HEHIQ scales 183
Table 47: Pearson's Correlations between 25-item WHHQ and three comparators 184
Table 48: Effect size and SRM of PROs N=105 (SF12v2 n=97) 186
Table 49: Comparison of content of CST conceptual framework v IN-CAM framework 218

List of equations

Equation 1: ICC formula 90
Equation 2: ICC absolute agreement 116
Equation 3: Standard error of measurement 117
Equation 4: SDC 121
Equation 5: SDC 2 122
Equation 6: Cohen's d 122

Abbreviations

ACQOL	Australian Centre on Quality of Life
AMED	Allied and Complementary Medicine Database
BDI	Beck Depression Index
CAM	Complementary and Alternative Medicine
CINAHL	Cumulative Index to Nursing and Allied Health Literature
CIMOS	Complementary and Integrative Medicine Outcome Scale
COSMIN	Consensus based Standards for the Selection of health Measurement Instruments
CST	Craniosacral Therapy (sometimes known as Cranio-Sacral Therapy)
CSTA	Craniosacral Therapy Association
CTT	Classical Test Theory
DAS	Disease Activity Score
EQ-5D	EuroQol – 5 Domain
FACIT	Functional Assessment of Chronic Illness Therapy
FDA	Food and Drug Administration
HADS	Hospital Anxiety and Depression Scale
HEHIQ	Harry Edwards Healing Impact Questionnaire
HIT-6	Headache Impact Test 6
HRQoL	Health Related Quality of Life
ICC	Intra-class Correlation Coefficient
ICF	International Classification of Functioning, disability and health
IHC	Integrated Health Care
IRT	Item Response Theory
IM	Integrative Medicine
MCS	Mental Component Summary
MIDAS	Migraine Disability Assessment Score
MOT	Medical Outcomes Trust
MYMOP	Measure Your Medical Outcome Profile
NCCIH	National Centre for Complementary and Integrative Health
NHS	National Health Service
OKS	Oxford Knee Score
OAB-V8	Overactive Bladder V8 Questionnaire
PANAS	Positive and Negative Affect Scale
PCS	Physical Component Summary
PGI-I	Patient Global Improvement Index
PGI	Patient Generated Index
PRISMA	Preferred Reported Items for Systematic Reviews and Meta-Analysis
PROs	Patient Reported Outcomes (Rest of World)
PROMs	Patient Reported Outcome Measures (UK)
PQSI	Pittsburgh Quality Sleep Index
PSQ	Perceived Stress Questionnaire
QALY	Quality of Adjusted Life Year
RCT	Randomized Controlled Trial
SAC	Self-Assessment of Change
SAC	Scientific Advisory Committee

SEIQoL	Schedule for Evaluation of Individual Quality of Life
SEM	Standard Error of Measurement
SF-12	Short Form 12
SF-36	Short Form 36
SRM	Standardized Response Mean
STAI	State Trait Anxiety Inventory
VAS	Visual Analogue Scale
WBQ-28	Wellbeing Questionnaire 28
WEMWBS	Warwick-Edinburgh Mental Wellbeing Scale
WHO	World Health Organisation
WHOQOL	World Health Organisations Quality of Life Assessment Tool

Declaration

This thesis is my own work and it has not been submitted for a degree at another university.

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree apart from the background material in the introduction section, pg. 27 and methods section pg. 99 which was previously submitted for a Master of Philosophy in Health Sciences degree in June 2012.

Research training

During my time of study at the University of Warwick I undertook the following research training:

October 2016	Concept elicitation for the development of clinical outcome assessments workshop, ISOQOL Conference, Copenhagen.
June 2016	Application of qualitative methods in complementary and alternative medicine. University of Warwick.
June 2016	Statistical methods for response shift detection, International Society for Quality of Life Webinar series (2 hours).
November 2015	Application of psychometrics for measuring health outcomes and quality of life. University of Sheffield.
May 2014	Mixed methods for health research. University of Warwick.
April 2014	Case report writing for complementary and alternative medicine. London South Bank University.
April 2014	Questionnaire design, application and data interpretation. University of Bristol.

Conference presentations

During my time at the University of Warwick I attended and presented my work at the following conferences.

- Complementary & Alternative Medicine Strategic Research and Development (CAMSTRAND) Conference, Southampton University, June 2017 (oral prize winner).
- World Congress Integrative Medicine and Health, Berlin, May 2017 (poster).
- International Society of Quality of Life Conference, Copenhagen, October 2016 (oral).
- Post Graduate Research Symposium, University of Warwick, June 2016 (oral).
- Research Council for Complementary Medicine (RCCM) Annual Conference, Middlesex University, November 2015 (poster).
- Complementary & Alternative Medicine Strategic Research and Development (CAMSTRAND) Conference, London Southbank University, June 2014 (poster).
- Post Graduate Research Symposium, University of Warwick, 2014 (poster).
- Complementary & Alternative Medicine Strategic Research and Development (CAMSTRAND) Conference, University of Westminster, July 2013 (oral).

Publications

Brough. N., Lindenmeyer. A., Thistlewaite, J., Lewith, G., Stewart-Brown, S. (2015) Perspectives on the effects and mechanisms of craniosacral therapy: A qualitative study of users' views. *European Journal of Integrative Medicine.* 7. 172-183.

Abstracts

Brough N., Parsons, H. and Stewart-Brown, S. (2017). Developing and evaluating a health related quality of life (HRQoL) questionnaire for craniosacral therapy (CST): Using qualitative methods to evaluate a conceptual framework*, BMC Complementary and Alternative Medicine.* 17(Suppl):P331.

Brough N., Parsons, H. and Stewart-Brown, S. (2016). The Warwick Holistic Health Questionnaire (WHHQ), developing and evaluating a patient reported outcome measure (PROM) for craniosacral therapy (CST): a mixed methods study. *Quality of Life Research*, 25: (1), p.69.

Brough, N., Parsons, H. and Stewart-Brown, S. (2015). Developing and evaluating a health related quality of life (HRQoL) questionnaire for craniosacral therapy (CST): Evaluating a conceptual framework. *European Journal of Integrative Medicine*, *6*(7), p.690.

Brough, N., Dolan, K., Harrison, H., Hemmons, J. and Low, C., 2015. Craniosacral Therapy Association (CSTA) survey of research skills & interests of members. *European Journal of Integrative Medicine*, *6*(7), pp.683-684.

Biggs, A., **Brough, N**., Cremer, C., Hatton, D., James, R., Kalinowska, L. and Maitland, F. (2015). A survey of craniosacral therapy (CST) practitioners: Profiling practice and informing organisational strategy. *European Journal of Integrative Medicine*, *6*(7), p.684.

Brough, N., Parsons, H. and Stewart-Brown, S.L. (2014). Developing and validating an outcome measure for craniosacral therapy: A mixed methods study. *European Journal of Integrative Medicine*, *6*(5), p.612.

Abstract

Aims

This thesis aims to design and evaluate a Patient Reported Outcome (PRO) capable of assessing change in Craniosacral Therapy (CST) users. CST is a mind-body based complementary therapy with limited evidence base partly due to lack of suitable PROs.

Methods

Mixed methods including focus groups and cognitive interviews were adopted to develop and evaluate a conceptual framework and the new PRO (Warwick Holistic Health Questionnaire WHHQ). Classical Test Theory and Exploratory Factor Analysis were used for psychometric testing.

Results

1. A conceptual framework (CF) of CST outcomes was refined and approved in 3 focus groups of practitioners and CST users.
2. 73 items were generated covering domains of the CF from an existing qualitative study of CST outcomes and PRO literature.
3. Face and content validity was tested in a consensus meeting with practitioners and two round of semi-structure interviews with CST users. The WHHQ was refined accordingly (52 items).
4. The WHHQ was pre-tested in cognitive interviews.
5. Item response, construct validity and item redundancy was assessed in 142 CST users.
6. The WHHQ was refined to 25 items including representations of new concepts in healthcare evaluation.
7. Reliability, internal consistency, external validity (SF-12v2, WEMWBS and HEHIQ), repeatability and responsiveness were assessed with 105 new CST users.

Conclusions

The conceptual framework of CST outcomes, the first of its kind, identifies important new domains of health and wellbeing including the development of self-awareness and the capacity to take responsibility for self. Measurement properties show the WHHQ is psychometrically sound, having good internal consistency and convergent validity with WEMWBS and HEHIQ. Test of repeatability showed mixed results: errors were bigger than the change value but comparable to WEMWBS and SF-12v2. Respondents reported improvements in health and wellbeing with small changes shown during evaluation of responsiveness. Testing in a larger sample might confirm these findings.

Acknowledgements

This journey has been immense for me and I am deeply grateful for the opportunity to undertake this study. I would like to express my sincere thanks to my supervisors Professor Sarah Stewart-Brown who skilfully managed me, knowing when to push and when to allow me the space I needed to digest and process challenging information or new learning. And to Dr Helen Parsons who has fostered my learning and understanding of quantitative methods patiently, despite my inadequacies at times. Other key personnel have been Sam Plumb for always been so positive and warm, Librarian, Samantha Johnson who has helped technically, and Dr Andrew Taylor for his attention when needed. Other academics who have supported me at times, the late Professor George Lewith (whose gift was making everyone feel valued) for his generosity of spirit, time and encouragement and Professor Nicky Robinson for her encouragement.

To my partner, Ian for his love and for holding the domestic fort and parenting our daughters Nesme and Neola when my mental and emotional attention was absent. To my parents, Sally and David, and Ian's parents Idona and Sidney, and his sister Viv for their love, support and for having gifted many hours and days of childcare. To Mandi Wong for the many extra school runs she's done in my absence.

To my dear friends, Helen Goodall, Margaret Baldwin, Carol Richards, Lisa Walmsley-Pitts, Janet Wright, Lesley Bailey, Maggie Bevington, Marie Tassone, Rachel Wilshaw, Monica Tomkins, Ruth Rowles, Su Menon, David Ellis and the late Vincent Winter for past love and support. To Penny Nisbet and Dr Harriet Senelane-Bhebe, Fiona Fazackerley and Dr Andrew Taylor for proof reading my work. To Amanda Biggs for always having the time to talk to me and holding the bigger vision when I was unable to; and Judy Hemmons for believing in me during my many crises of confidence and holding my vulnerability.

To the Warwick University Chancellor's Scholarship fund and the Craniosacral Therapy Association for funding this work; the CST practitioners who supported me both practically and with encouragement along the way. And finally, to the participants in this study for their time and depth of sharing.

Dedication

This thesis has been written to fulfil part of the examination criteria for a Doctor of Philosophy Degree; but in addition, I (NB) am a craniosacral therapy (CST) practitioner and have been in practice since graduating from my CST training in 2006. The motivation for this work stems from the positive changes I see in clients within my practice having used CST to support them, alongside the lack of rigorous research studies that show CST's effectiveness. A lack of suitable patient reported outcome measures (PROs) may have contributed to the dearth of literature. This thesis is my contribution to the field of CST and it is hoped that CST clients and practitioners will benefit from any outcomes resulting from this thesis.

1. Introduction

1.1 Chapter Overview

In this section the aims and objectives are presented and a background to the study is given. Complementary and alternative medicine (CAM) is defined and the ontological stances of CAM are introduced. CST and outcomes of importance to CST users are introduced. An overview of outcome measurement in health is given and conceptual frameworks from a bio-medicine perspective and CAM are presented. The concepts of spiritual wellbeing and health-related quality of life are introduced. The section concludes by looking at measuring outcomes of CST practice and different types of outcome measures are briefly discussed to evaluate their suitability to assess CST.

1.1.1 Aims and objectives

The study aim is to:

- To identify and evaluate PROs suitable for evaluating outcomes of CST. If none found, to develop a patient reported outcome measure suitable for use in CST clinical practice.

The study objectives as set out below:

1. To develop a conceptual framework of CST outcomes which includes CST users' views.
2. To identify a set of candidate patient reported outcomes (PROs) of CST which are used within the literature.
3. To assess the candidate CST PROs against the conceptual framework.
4. To validate the candidate PROs for use in CST, (or)
5. To develop and evaluate a new patient reported outcome measure to assess changes in health and wellbeing of CST users which met the following criteria:
    - to be patient reported
    - be easy to complete, score and interpret
    - for use in clinical practice

- Include both CST users and CST practitioners in the development and evaluation process.

PRO development is an iterative process and each aim informed the next stage of conceptualisation, evaluation and validation.

6. Report the findings of the above process in a thesis.

1.1.2 Background

Complementary and alternative medicines (CAM) are increasing in popularity in the UK (Eardley, *et al*., 2012). In England, the lifetime and 12-month use of CAM is now 44% and 26% respectively (Hunt *et al*., 2010). Due in part to patient demand, some CAM treatments are now provided within the National Health Service (NHS) and GPs are using CAM as part of their health improvement programs (NHSTA, 2011). The evidence base for many of these therapies however, remains weak.

1.1.3 Defining CAM

According to the United States National Centre for Complementary and Integrative Health (NCCIH) (NCCIH. 2015) people often use the words “alternative” and “complementary” interchangeably; but the two terms refer to different concepts:

- “Complementary” generally refers to using a non-mainstream approach **together with** *conventional medicine*.
- “Alternative” refers to using a non-mainstream approach **in place of** conventional medicine.

Zollman's (1999) definition provides a good example of CAM: *"Complementary and Alternative Medicine (CAM) is a broad domain of healing resources that encompasses all health systems, modalities, and practices and their accompanying theories and beliefs, other than those intrinsic to the politically dominant health system of a particular society or culture in a given historical period.*" True alternative medicine is not common. Most people use non-mainstream approaches along with conventional treatments and the boundaries between complementary and conventional medicine overlap and change with time.

The term Integrative Medicine (IM) is also used. Bell *et al*., suggest that "*Integrative Medicine (IM) is a system of care that emphasises wellness and healing of the entire person (bio-psycho-socio-spiritual dimensions) as primary goals, drawing on both conventional and CAM approaches in the context of a supportive and effective physician-patient relationship"* (2002, p.133). Both IM and CAM propose that the origins of 'dis-ease' are multi-factorial with factors operating together in a complex matrix where all items are as relevant as each other. The factors include genetic, environmental, social, physical, emotional, psychological and spiritual issues (Bell *et al*., 2002, p.136.). This complexity of disease aetiology and thus pathways to health has implications for the selection of approaches to instrument validation and measurement theory to be discussed later in this thesis.

1.1.4 Concepts of health: a complementary and integrative medicine perspective

The main ontological stances of CAM and Integrative Medicine (IM) are presented here, as these are important to orientate the reader to the intentions of this thesis. Many systems of CAM, e.g. craniosacral therapy (CST), classical homeopathy, traditional Chinese medicine and Ayurvedic medicine hold the view that a disease may manifest at the emotional, mental or spiritual level as well as on the physical plane (Jonas, 1999). These CAM systems have 'vibrational medicine' (Gerber, 1988) as an underlying mechanism and espouse the belief that the physical body is a complex network of interwoven energy fields representing the physical and cellular framework and nourished by 'subtle energetic systems' that coordinate the life force in the body. To illustrate these concepts in the context of the traditional Indian medicine approach (Ayurveda), the ancient literature of the Vedas refers to these subtle energy systems as the chakras (or wheels), described as spinning spheres of bio-energetic activity emanating from the major nerve ganglia in front of the spinal column (Judith, 1996). The chakras are held to be specialized energy transformers which take subtle energy and distribute it to the major glands, nerve centres, and organs of the body. They are also held to be affected by emotional and spiritual states so that when an individual has unresolved emotional issues in any of these areas, chakra dysfunction may occur and lead to deprivation of

nutritive energy flows to that region of the body. Cellular imbalance and physical disease will ultimately occur if the chakra blockage is chronic (Gerber, 1988). It is apparent therefore that in the Ayurvedic system, as in other CAM, an effective intervention may need to address spiritual and emotional distress to relieve 'dis-ease' on a physical level.

Although concepts of health and disease common in the CAM and IM fields are regarded by many as outlandish and without foundation; enquiry and research in the fields of neuroscience and neuro-physiology (Hariri *et al*., 2000; Frith and Frith, 2003; Solomon and Siegel, 2003; (Edwards, 2015) is providing a scientific basis for body energy systems.

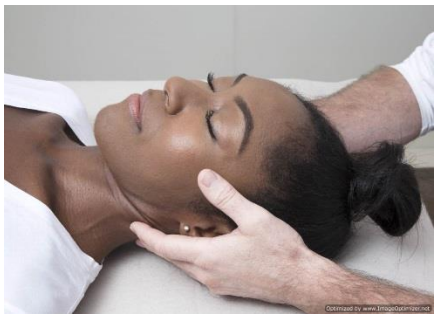
1.1.5 Craniosacral Therapy (CST)

CST is a CAM developed from clinical experience within the field of osteopathy (Sutherland, 1990). It is a 'hands on' therapy which is thought to assist the body's natural capacity to self-repair. The practice is not currently regulated and practitioners may choose to become a member of an association if they wish, for example the Craniosacral Therapy Association (CSTA)(www.craniosacral.co.uk). CSTA members have all undertaken a standardised one or two year/s' training at an accredited school; they adhere to the CSTA code of ethics, participate in regular continuing professional development and hold professional indemnity insurance.

Having a session of CST involves lying fully clothed on a treatment table, then the practitioner makes light contact on the client's body with their hands (see Figure 1). It involves a therapeutic process between client and practitioner which emerges from the 'intention' of the client and practitioner to enter a space from which the potential for change may arise. The intentions may be determined through consultation with the client prior to the hands-on work commencing. It is not uncommon within practice for a CST client to present with a physical problem for which a psycho/emotional trauma is the origin. These processes are skilfully held and facilitated with compassion and presence. The mindset of CST practitioners and their beliefs about health, illness and bodily functions differ from those of

conventional medical practitioners. Skills such as “presence” and “intentionality” are used as a foundation from which to engage with the mind-body-spiritual aspects of an individual coming for CST.

Figure 1: CST session vault hold (image used with permission)



Practitioners use their all their senses to ‘listen’ and observe the client’s body and are trained to feel congestion or restriction in the movement of energy in the various systems of the body, for example; the heart, blood flow and cerebral spinal fluid in the CST system. The practitioner uses their hands, and an intention to amplify any abnormal patterns, which allows the body to better sense this and return to an easier way of functioning.

People present for CST for diverse reasons. Brough (2012) reported that participants in her study (n=29) engaged in CST for various reasons including physical and mental health problems, being in pursuit of greater wellbeing particularly psychological and spiritual wellbeing. Some sought CST as an alternative to conventional medicine particularly when they had experienced poor care or care which had not met their needs. Almost all participants presented with health problems which warranted care within the NHS; most had some experience of NHS care already and were seeking to improve their health beyond what the NHS had been able to offer. Participants reported that they were looking for: relaxation and stress relief; support with rehabilitation; a holistic approach to healthcare; an alternative to conventional medication or more invasive techniques as motivations for having CST. The CSTA carried out a survey (2013) to identify why people presented for CST and similar reasons were identified. Practitioners reported that

people came to combat stress related illnesses, to address physical pain and to support wellbeing.

The hands-on component of a session can take up to 40 minutes to one hour. Sessions typically cost between £30 and £50 in most of the UK, although costs can be higher in London. Treatment plans are negotiated according to client's needs, presenting symptoms and how their bodies respond to treatment.

CST is popular with clients who report important changes to their health (Brough *et al.*, 2015) but like with many other CAM approaches, the current evidence base is limited. One of the key barriers to achieving an evidence base is the identification of a valid outcome measure. To identify appropriate measures, it is important to establish patients' perspectives on health outcomes from CST.

One such study was carried out. Brough (2012) carried out twenty-nine semi-structured interviews with participants who had attended six or more sessions of CST within a 12-month time frame using the following topic guide:

- Why did they come for CST?
- What was their experience of CST?
- Asked to summarise the important outcomes or changes they had noticed whilst having CST.
- How did they feel after a session?
- For how long did those feelings last?
- To mention anything else that they felt was important in relation to CST that had not already been discussed.

Inductive thematic analysis was used to identify the outcomes from the users' perspective, analyse and report on the data. Findings showed that health gains from CST are usually holistic, that is they occur in more than one of the three domains of body, mind and spirit. Bodily symptoms in which participants reported change included pain, limitations in functioning and mobility. Positive changes relating to the mind were reflected by a heightened self-awareness which developed in six areas:

- more awareness of psycho-emotional aspects of self
- changes in self-concept
- understanding mind-body-spirit links
- improvements in interpersonal relationships
- better coping strategies
- enhanced engagement of self-care and capacity to manage health problems

Spiritual changes included a sense of connectedness with self, others, the wider universe and a general sense of enhanced wellbeing.

These findings of holistic effects and a focus on the positive suggest that wellbeing could potentially be a good basis for a CST outcome measure. They also suggest the need for a measure to consider self-awareness as an essential part of health and that spiritual wellbeing needs to be covered by the components also. These findings form the criteria which a questionnaire will need to fulfill to capture all of the outcomes of importance to CST users, see Table 1.

Table 1: Domains and outcomes of importance for CST

Physical Wellbeing, Mental Wellbeing, Spiritual Wellbeing, Social Wellbeing.
Domains to include aspects of self- awareness, the mind-body-spirit link, interpersonal relationships and engagement in self-care.

In the field of CST, the evidence base is currently limited and having a tool to use in clinical practice which includes items that are patient reported and is acceptable to practitioners, will lay a solid foundation for future and ongoing research processes. Evaluating practice by assessing treatment outcomes is an important way for practitioners to increase the effectiveness of their work, contribute to data collection initiatives and in turn develop the evidence base.

1.1.6 Concepts of health from the bio-medicine perspective

As the authority on health amongst the United Nations states, the World Health Organisation (WHO) has defined and conceptualised health. The WHO's original definition of health *"a state of complete physical, mental and social wellbeing not merely the absence of disease"* (WHO, 1946) mentions three domains now widely used in health measurement.

As the introduction so far has explained, a holistic perspective includes the domains of physiology, mental, physical, social and spiritual wellbeing which are all intricately linked within an individual and in the wider context of society. Wellbeing is thus intrinsically holistic and in the context of this thesis is presented as the positive end of the health spectrum which runs from illness to wellbeing. The reductionist bio-medical approach separates these domains, focusing on physical and mental disease separately and on pathophysiology, but often and confusingly referring to these as mental and physical *health*. Biomedical outcome measures thus tend to be reductionist i.e. not holistic and favour objective measurement over measures which capture patient and client experience.

In regard to the WHO having defined health as wellbeing, this has created confusion as the word 'health' continues to be associated to health services and disease. If health is defined as wellbeing, health cannot be the spectrum. It is not helpful to propose this, it may be better to use wellbeing as the positive end and illness and the negative pole. Wellbeing is intrinsically holistic. The concept of wellbeing is examined further in the discussion chapter (pg. 211). In the context of CST, the domains of interest are mental, physical, social and spiritual wellbeing captured through self-report or patient experience.

To orientate the reader, the concept of 'mental wellbeing' is used next as an example to provide a brief overview of the inter-disciplinary and historical context.

1.1.7 Inter-disciplinary and historical context

The concept of 'mental wellbeing' sits as one of the domains of wellbeing in broader context of 'health' defined here as the spectrum from illness to wellbeing. In addition to those working in health and social-care, different disciplines are interested in mental wellbeing including psychology, philosophy, social science and economics (2015) and all have different needs and perspectives on measuring wellbeing (Stewart-Brown, Forthcoming 2017).

Social scientists and economists use measures of happiness and life satisfaction in their studies (S., 2015; Stewart-Brown, 2015b) and make distinctions between subjective and objective wellbeing which focus on the social circumstances and structures they regard as necessary for wellbeing. Psychologists focus on functioning or behaviour rather than feelings and define different components of mental wellbeing (Ryff, 1989) developing multidimensional measurement scales to score the different attributes they recognise. Philosophers and spiritual leaders' aims are to help individuals to develop themselves, more akin with the psychologists' concepts of functioning well encouraging the cultivation of character traits and behaviours that enhance happiness for self and others.
In the context of this thesis the definition of mental wellbeing which has been adopted is the public health one which encompasses both feeling good and functioning well.

The concepts of physical wellbeing, social wellbeing and spiritual wellbeing are also of interest to many different disciplines with different origins. In the context of this thesis the concept of physical wellbeing 'is a spectrum represented by physical symptoms that may influence posture, inhibit function and mobility. With optimum physical functioning at one end of the continuum with ability and daily activities and how they are affected (poor physical function) at the other end of the continuum' (Brough N., *et al*. 2014). social wellbeing means 'a sense of relatedness and connectedness to other people (NEF, 2009)'. Spiritual wellbeing is explained on

pg. 38. The way in which they are envisaged to link together in the context of health measurement for CST is explored in the discussion chapter, pg.210.

1.2 Outcome measurement in health

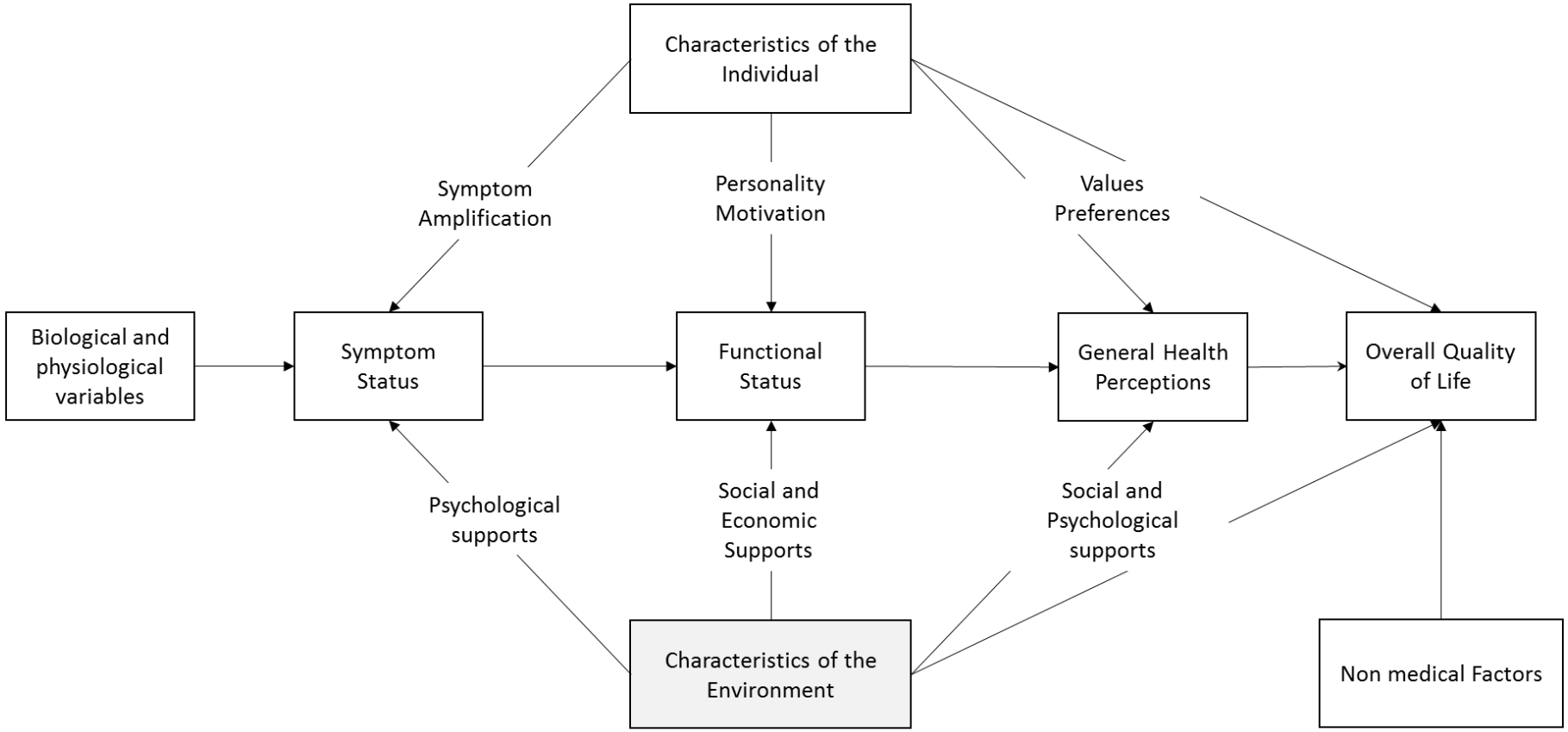
A shift in thinking has seen concepts such as health related quality of life (HRQoL) and mental wellbeing come to prominence in health outcome measurement (Bowling, 2005). Both are based on subjective assessment of health states by patients. This subjective assessment has contributed to the development of a more sophisticated understanding of health. In the past, the latter has been primarily defined as the absence of specific diseases whose presence can only be determined by the medical profession. In the 21st century, health is much more broadly defined and health outcome measures have come to reflect these broader definitions.

For example, the WHO International Classification of Functioning, Disability and Health (ICF) which is a conceptual framework which is now the dominant conceptual basis for the definition, measurement and policy formulations for health and disability (WHO, 2002). It reflects the shift of emphasis in health care from acute life-threatening conditions to chronic illness and the accompanying shift in emphasis of outcome measures from rates of mortality (Wilkin *et al.*, 1992) to the assessment of functioning at the level of the whole human being, in day-to-day life (WHO, 2002) which is encompassed in the concept of Health Related Quality of Life (HRQoL) and is discussed in more depth on pg. 38.

Wilson and Cleary (1995) presented a conceptual model for measuring HRQoL (see Figure 2). The areas of the model, range from the molecular and cellular to the impact of health or disease on individuals in their environment and quality of life. The model also hypothesises causal pathways of how factors can influence HRQoL. The arrows in the model show the direction of influence but there may be shared relationships. Objective measures such as biological and physiological variables which can be obtained by direct measurement are mainly found on the left-hand side of the Wilson and Cleary model; and symptoms and functional status which are subjective and have non-observable characteristics such as pain and fatigue are positioned in the centre of the model. Health perception and QoL are non-

observable constructs and are located on the right of the model. The non-observable characteristics are referred to as 'constructs, latent traits or factors' by psychologists (Fayers & Machin, 2007) and measurement theories of how to measure non-observed characteristics indirectly were born. Measurement theories will be discussed later in this thesis but to conclude this section, multi-item measurement questionnaires are a result of measuring observable characteristics related to non-observable constructs. When using multi-item questionnaires it is important to understand the relationship between the items and the construct to be measured (De Vet *et al.*, 2011). This underlying relationship is depicted in a conceptual framework, which then determines the measurement theory to be used in developing new questionnaires.

Figure 2: Wilson and Cleary conceptual model for health-related quality of life



Relationships among measures of patient outcomes in a health-related quality of life conceptual model. (Wilson & Cleary, 1995) pg. 60

35

1.2.1 Conceptual frameworks: reflective and formative models

As in the Wilson and Cleary model of HRQoL (1995) above a conceptual framework presents hypothesised relationships of how the constructs and items of a questionnaire are linked. The aim of numerical methods in QoL research according to Fayers and Machin (2007) *"is to test the adequacy and validity of models based upon the postulated constructs, and estimation of the values of the latent variables that comprise those models*" (pg. 3). To determine the model type an example of anxiety and life stress (constructs to be measured) will be used: if anxiety is manifest in the items presented on the questionnaire such as 'worrying thoughts', 'panic' and 'restlessness' this is known as reflective model; if the construct (life stress) is a result of the presented items, this is known as a formative model as the items or 'causal indicators' (Fayers & Hand, 1997; Fayers & Machin, 2007). These examples form the construct items 'job loss', 'death in the family' and 'divorce' equate to 'life stress', taken from De vet *et al*., (2011). The distinction between the two models is not always straight forward and concepts such as HRQoL combine elements of both. The relevance of introducing reflective and formative models is that conceptual frameworks for CAM are in their infancy and it is not yet clear if these models are appropriate in this area of work.

1.2.2 Conceptual framework for CAM

As CAM is a relatively new area of research, outcome measures tend to be adopted from other areas of healthcare such as disease specific and HRQoL instruments that quantify change over a period of time in specific symptoms. Standardised measures that assess generic health related quality of life may be used, but do not cover all aspects of CAM (Verhoef, 2006). As a result of the mismatch between available measures and the therapies under study CAM researchers (Bell *et al*., 2002; Ng, 2005; Verhoef *et al*., 2006; Ritenbaugh *et al*., 2011; Draper, 2012) have started to develop explicit conceptual frameworks which depict the basis of CAM and IM and throw light on the gaps in available measures. One such framework is evident in the IN-CAM outcomes database (Verhoef et al., 2006).

1.2.3 IN-CAM outcomes database

IN-CAM is an internationally recognized CAM research network. Members (CAM researchers, practitioners and students) have developed an online database of outcome measures of importance to CAM and Integrative Health Care (IHC), effectiveness and efficacy research. The conceptual framework shown in Figure 3 underpins the IN-CAM database and has nine domains (Verhoef et al., 2006) the context of healing; the process of healing; holistic outcomes; health related quality of life outcomes; spiritual outcomes; psychological outcomes; physical outcomes; social outcomes; and individualised measures. Many CAM interventions are conceptualized as whole systems of care, made up of multiple components that provide an individualized approach to treatment (Verhoef *et al.*, 2005a).

Figure 3: Framework of health outcome domains relevant to CAM research and practice

Taken from http://www.iscmr.org/content/welcome-cam-health-outcomes-database [accessed 24 August 2017].

The IN-CAM website includes practical information on outcome measures within the framework of domains that makes them easily accessible, and can assist

researchers and practitioners in framing their approach (research or clinical) within a whole person perspective and/or a wellness orientation.

In the light of the above, concepts of both HRQoL and spiritual wellbeing are important in the context of CAM; as CAM practitioners usually adopt a holistic approach that considers the whole person in their environment and incorporates multiple interventions which work together to synergistically improve health (Deng, Weber, Sood & Kemper, 2010, cited in Hunter, 2013).

1.2.4 Health related quality of life

The concept of HRQoL was developed in response to concerns about the inadequacy of traditional medical outcome measures. Philosophically, HRQoL is a broad ranging concept which can be affected by many different and connected factors: the person's physical health, psychological state, level of independence, social relationships, personal beliefs and their relationship to salient features of their environment. Measures of HRQoL usually include physical, emotional and social dimensions of health and functioning, for example the Short-Form 36 item Health Survey (SF-36) (Ware & Sherbourne, 1992) which is a measure of health status that includes work and role performance. At first glance the concepts appear to capture similar domains as those mentioned in the section Conceptual framework for CAM and will be discussed in more depth later in this thesis (pg. 216).

1.2.5 Spiritual wellbeing

Holistically, assessing health requires that outcome measures capture the health domains which encompass the whole person, i.e. the domains of mind, body (or physical health), spirituality and social health (Verhoef, 2006; Luff & Thomas, 2000). Whilst the first two domains are familiar within healthcare, the importance and nature of the concept of spirituality within healthcare is debated (Ben-Arye, 2008).

Outcomes associated with spirituality are commonly documented within CAM literature (Verhoef, 2006; Eton, 2005; Bell, 2002; Bell, 2004 and Ritenbaugh, 2011)

and include attributes such as awareness, enablement, transformation, hope, peace, relaxation, and balance. In areas of healthcare where CAM use is common such as in oncology (Ben-Arye, 2008), management of chronic illnesses (Paterson & Britten, 2003), musculoskeletal disorders (Hsu, 2010) and mental health. Spirituality is important as *"individuals see spirituality as an essential part of mental wellbeing"* (Gilbert & Parkes, 2011. p.3). In the context of this thesis spiritual wellbeing includes 'personal and internally focused elements such as core values, philosophy and meaning of life or the way one conducts life. It is the search for the sacred or divine through any life experience (Mytko & Knight, 1999) and connectedness to self, nature and the wider universe'.

1.2.6 The relationship between HRQoL and wellbeing

HRQoL domains measure people's health status, therefore in the WHO definition their wellbeing. Johnston et al., (2013) presented a definition of wellbeing as *"subjective bodily and emotional states; how an individual feels; a state of mind distinct from functioning that pertains to behaviours and activities."* Or as defined by the Cambridge English Dictionary *"a state of feeling healthy and happy"* and adopted by social scientists as meaning a transient state of mind, often determined by circumstances beyond the control of the individual.

The WHO (2001) defined mental health as "*a state of wellbeing in which every individual realises his or her own potential can cope with the normal stresses of life can work productively and fruitfully and is able to make a contribution to his or her community."* According to NEF, an independent think tank (whose partners include Cambridge University) (NEF, 2009) the science of 'subjective wellbeing' suggests that as well as experiencing good feelings, people need:

- A sense of individual vitality.
- To undertake activities which are meaningful, engaging, and which make them feel competent and autonomous.

- A stock of inner resources to help them cope when things go wrong and be resilient to changes beyond their immediate control.

It is also crucial that people feel a sense of relatedness to other people, so that in addition to the personal, internally focused elements, people's social experiences – the degree to which they have supportive relationships and a sense of connection with others – form a vital aspect of wellbeing. To summarise, wellbeing is experienced when people's lives are going well; although the importance of the elements of wellbeing mentioned above might change as an individual's circumstances change.

As the nature of wellbeing and HRQoL are subjective regarding health measurement, here lies one of its challenges: people's assessment of their health state varies not only when one of the main domains linked to HRQoL changes; but also with the extent to which they have adapted to their illness and this can and does change over time. Called 'response shifts' by Schwartz and Sprangers (1999), these shifts in internal standards (recalibration), values (reprioritization) and conceptualisation (reconceptualization) are a valuable strategy for coping with the reality of a chronic disease, by adjusting one's expectations during ever changing health states. Response shift has implications for health measurement in CST which is restricted in discussion in this work.

1.2.7 Measuring outcomes for CST

It was important in this study to check if the IN-CAM framework was appropriate for CST. To do this, data from Brough's (2012) study was used to develop the framework of CST outcomes which will be evaluated as part of this study. Brough *et al*. (2015) hypothesised that awareness changes how individuals relate to self, others and their environment and therefore how they respond to self-report PROs. If awareness is a medium of change and not a component of a domain, as presented in the literature (Verhoef *et al*., 2006), it is essential to be able to measure any shifts in awareness in all domains and its components both simultaneously and independently. So, whilst individuals will present with their

own unique history and biography, it is necessary to reinforce the point that all domains and their components are as relevant as each other in the journey of regaining health and wellbeing. And individuals will need to develop awareness to realise that 'awareness' was not there initially. Therefore, it is important that the correct conceptual framework and subsequently correct outcome measure be used to match the healthcare modality and that the outcome measure is thus fit for purpose.

1.3 Typology of Measures

In this section the different types of outcome measures are briefly discussed and evaluated for their suitability to assess CST. Examples of leading types of measures are provided. At a macro level measures can be defined as generic, utility, specific or individualised (Garratt *et al*., 2002).

1.3.1 Generic measures

Generic measures are used for a broad range of health problems and due to their broad scope can be used to make comparisons across therapies and can be sub-defined as profile measures e.g. the Short Form - 36 (Ware & Sherbourne, 1992; Hadorn *et al*., 1995). The term profile is used for a "*multidimensional construct that consists of different dimensions for which a score is presented for each dimension"* (de Vet *et al*., 2011, p.51). The disadvantages of generic measures are the inability to detect specific details in a specific context e.g. CST as an intervention (unless the outcome measure was designed specifically for measuring CST) or disease specific symptoms.

1.3.2 Utility measures

Utility measures have been developed from economic and decision theory to identify the health states in a population as well as individual patients at a specific point in time (Fitzpatrick *et al*, cite Drummond, (1993) and Bakker and Van der Linden (1995). The EuroQol EQ-5D (Euroqol Group, 1990), is the most common utility measure of disease burden in the UK and has been validated in many different patient populations worldwide. It is used to calculate the quality of

adjusted life year (QALY) which has an important role to play in guiding policy and practice. Utility measures may be relevant for health economic assessment of CST against other therapies in due course, but are unsuitable for effectiveness research as important health consequences are excluded.

1.3.3 Specific measures

Specific measures can include: conditions or disease measures such as the Disease Activity Score (DAS) for rheumatoid arthritis (van der Heijde, 1990); site specific measures e.g. the Oxford Knee Score (OKS) Dawson *et al*., (1998); domain specific measures for example: fatigue; population specific measures e.g. children; or intervention-specific measures, for example the Oxford Hip Score (Dawson *et al*., 1996) for hip replacement surgery. Disease and site-specific measures are inappropriate for the evaluation of a holistic therapy like CST; as adopting them would contradict the CAM philosophy of working holistically and not with simply one part of a person.

1.3.4 Individualised measures

Patient-centred measures of health aim to be more sensitive to the individual's needs, demands and change in status e.g. the Patient Generated Index (PGI) (Ruta *et al*., 1994) offers a patient centred approach to the evaluation of disease specific health related quality of life. The Schedule for Evaluation of Individual Quality of Life (SEIQoL) (O'Boyle, 1993) allows the respondents to nominate the 'top five' domains relevant to their quality of life. The Measure Your Medical Outcome Profile (MYMOPv2) (Paterson, 1996) was developed using a CAM population and is discussed in more depth in the literature review section (pg. 57). The individualised nature of these measures means there can be a lack of comparable evidence.

1.3.5 Patient reported outcomes

This section defines Patient Reported Outcomes (PROs) and demonstrates how they are applied in a clinical setting as part of the NHS Outcomes Framework (DOH, 2010).

PROs are a means of collecting information as perceived by patients themselves. The Food and Drug Administration (FDA) define patient-reported outcome as *"a measurement based on a report that comes directly from the patient about the status of a patient's health condition without amendment or interpretation of the patient's response by a clinician or anyone else. A PRO can be measured by self-report or by interview provided that the interviewer records only the patient's response."* (2009, pg.32)

PROs aim to provide an objective measure of a subjective construct: that is, an individual's experiences and concerns in relation to their health, health care and quality of life (Fitzpatrick *et al*., 1998, Ganz, 2002, cited in Haywood, 2006, pg. 189) they are questionnaires which contain numerous questions or items to measure outcome variables.

Patient-reported outcome measures (PROMS) as they are known in the UK, are applied in a clinical setting as part of the NHS Outcomes Framework in England (DOH, 2010). PROMS data can be used to assess the clinical quality of providers; research what works; evaluate the effectiveness and cost-effectiveness of different technical approaches to care; assess the relative health status of patients before operations; and support the reduction of health inequalities (HSCIC, 2013).

1.3.6 Chapter summary

This section provided the background to this thesis. It introduced and defined CAM. It introduced CST and outcomes of importance to CST users. It presented the way in which health is conceptualised in both the bio-medical model and CAM models of health and introduced conceptual frameworks for such models. The concepts of health-related quality of Life (HRQoL) and spiritual wellbeing were introduced and the relationship between HRQoL and wellbeing were discussed. Health outcome measurement was introduced, the typologies of measures were presented and PRO was defined.

2 Literature Review

This section describes a series of literature reviews designed to establish if the intended research had already been undertaken; to identify and appraise PROs that are currently being used to evaluate CST; and to identify and appraise PROs that had been developed for use in CAM to establish whether any of these are suitable for use in CST. It presents the criteria that a CST PRO needs to fulfil. The search strategies, inclusion and exclusion criteria for each review are shown. The results of the searches are reported and eligible studies are reviewed. The section concludes by reporting on the methodological challenges associated with literature searching for PROs developed for CST/CAM and a summary of the results of the review.

2.1 Introduction to literature review

Although personal knowledge of the field and conversations with experts suggested that no PRO had been developed for CST, it was important to confirm this with a systematic search.

It was also important to establish that no existing PROs developed for other purposes could adequately cover the outcomes of CST. In section one (pg.23) these were described as covering the domains of: mind, body and spirit; heightened self-awareness, particularly of psycho-emotional aspects of self and mind-body-spirit links; enhanced engagement of self-care and capacity to manage health problems and interpersonal relationships.

Because the PRO literature is extensive and approaches to searching are not as well established as they are for intervention studies, a pragmatic approach was adopted; including searching for evaluation studies of CST and identifying PROs that had been used for this purpose; searching for PROs which had been developed for use in CAM more generally; including reviews of suitable PROs and finally consulting with experts in the field.

2.2 Inclusion criteria for a CST PRO

Table 2 below lists the questionnaire inclusion criteria including outcomes of importance to CST users, and criteria relating to practical issues such as cost, the use of English Language and length. It was hypothesised that no one questionnaire would cover all domains necessary, so if a questionnaire met some criteria, for example, had items that matched the mental domain, it would be considered.

Table 2: Criteria for selecting questionnaires

Inclusion criteria
Domains to include mental wellbeing, spiritual wellbeing, physical functioning, social wellbeing
Has items that capture heightened self-awareness
Has items that show enhanced capacity for self-care/taking responsibility for one's health
Designed for CST/CAM and/or used in CST/CAM settings
Free to use
English only
< 40 items covering all mentioned domains

2.3 Systematic search to identify PROs for CST

2.3.1 Search 1: identifying PROs for CST

The aim of this search was to establish whether any PROs had been developed specifically for use in CST.

Methods for search 1

Where applicable, the PRISMA guidelines for systematic literature reviews were used (Moher *et al.*, 2009). A systematic search was undertaken in September 2014 and updated in April 2017. Published papers were identified using general medical databases and subject specific databases (Medline and EMBASE databases via OVIDSP; Allied and Complementary Medicine Database (AMED); Cumulative Index

to Nursing and Allied Health Literature (CINAHL) via EBSCOhost; PsycINFO via ProQuest and Science Direct) in the first instance. References quoted in retrieved articles were examined to ensure articles known to researchers in the field had not been missed. Search limits were applied and included: English language articles, PROs for adults (aged ≥ 18) and published in peer review journals. Details of the search are shown in Table 3 below. An example of the search strategy used in all databases named above is shown in Table 4

Table 3: Search 1 search terms

Search terms:	Outcome Assessment (Health Care)", "Surveys or Questionnaires", Outcomes database*, patient reported outcome measure*, patient reported outcome questionnaire*, "patient-reported outcome*" "primary care outcome measure", "Craniosacral therap*, cranio-sacral therap*,

Table 4: Search 1 search strategies and results – Medline

Search No.	Search strategy Medline	# Results
1	Exp "Outcome Assessment (Health Care)"/	62,945
2	Exp "surveys or questionnaires"	917,180
3	Exp "Surveys and Questionnaires"/	868,214
4	(patient reported outcome measure* or primary care outcome measure* or patient reported outcome questionnaire*).mp.	2,139
5	outcomes database*.mp. or Treatment Outcome/	829,290
6	PROM*.mp	131,7902
7	#1 OR #2 OR #3 OR #4 OR #5 OR #6	304,2077
8	craniosacral therap* or cranio-sacral therap*).mp.	59
9	CST.mp	2,706
10	#8 OR #9	2,756
11	#9 AND #10	638
12	#11 AND #8	25

The following inclusion and exclusion criteria were applied to the list of articles found to see if anyone had developed a CST PRO.

Inclusion criteria

i. Papers describing the development of a PRO for the evaluation of CST.

Exclusion Criteria

i. Not published in English language
ii. Not published in a peer reviewed journal
iii. Developed or evaluated for patients > age of 18 years

2.3.2 Results of search 1

As shown in Table 4, 25 articles were found. All were excluded as they did not meet the inclusion criteria. This search yielded no papers describing the development of a PRO for CST. This suggests that a therapy specific patient reported outcome measure for CST does not exist. Considering these results, a second systematic search to identify and appraise PRO measures which have been used in CST evaluation studies.

2.4 Search 2: identifying CST studies using PROs for evaluation

The aim of search 2 was to identify PROs applied in studies of CST.

2.4.1 Methods for search 2

Examples of the search terms used can be seen in Table 5.

Table 5: Search 2 identifying PROs used in CST studies

Search terms:	Outcome Assessment (Health Care)", "Surveys and Questionnaires", patient reported outcome measure*, patient reported outcome questionnaire*, "patient-reported outcomes and quality of life questionnaire*", CST.mp, "Craniosacral therap*, or cranio-sacral*"

Electronic databases (Medline, AMED and EMBASE databases via OVIDSP; CINAHL via EBSCOhost; PSYCHTESTS via ProQuest and Science direct) were searched for published CST papers. An example of the search strategy used in Medline is seen below.

Table 6 Search results identifying CST studies using PROs for evaluation

Search No.	Search strategy Medline	# Results
1	Exp “Outcome Assessment (Health Care)"/	62,945
2	Exp “surveys or questionnaires”	917,180
3	Exp "Surveys and Questionnaires"/	868,214
4	(patient reported outcome measure* or primary care outcome measure* or patient reported outcome questionnaire*).mp.	2,139
5	outcomes database*.mp. or Treatment Outcome/	829,290
6	PROM*.mp	131,7902
7	"Quality of Life"/ or quality of life questionnaire*.mp.	330,0520
8	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7	295,3486
9	craniosacral therap* or cranio-sacral therap*).mp.	59
10	CST.mp	2,706
11	#8 OR #9	2,756
12	#8 AND #10	638
13	#11 AND #8	25

The following inclusion and exclusion criteria were applied to articles found:

Inclusion criteria

- Studies evaluating the effects of CST used as sole therapy
- In any disease or condition
- Using any quantitative methodology

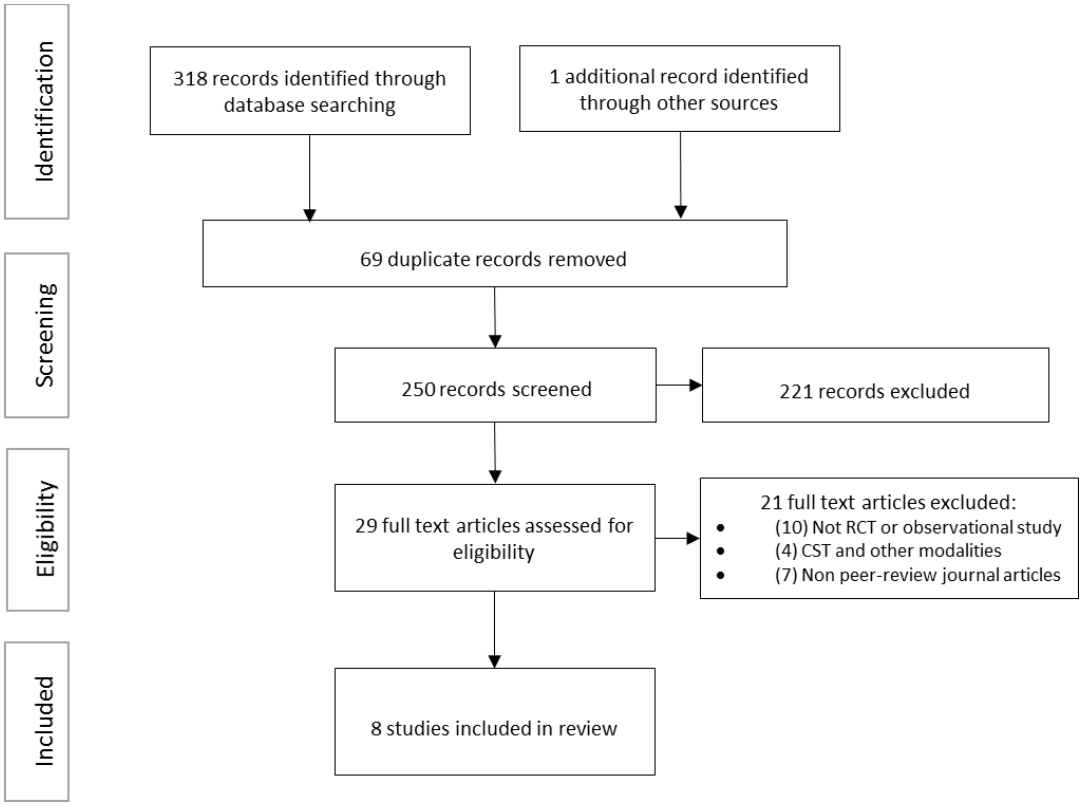
Exclusion Criteria

- Participants younger than 18 years
- Not published in a peer-review journal
- Not published in English
- Studies in which the effects or CST were evaluated combined with another modality (e.g. acupuncture).

2.4.2 Results of search 2

Figure 4 shows the number of records identified, screened, retrieved to assess for eligibility and those studies included in the review.

Figure 4: Results of the systematic search for PROs used in studies evaluating the effects of CST



A total of 319 records were identified by the search. During the screening process, 69 duplicates were removed leaving 250 abstracts to be screened against the inclusion and exclusion criteria for relevance. 221 records were excluded as they did not meet with the inclusion criteria. 29 full text articles were then accessed for eligibility: 7 were excluded as they were not full peer-review journal articles; 4 were excluded as CST and other modalities were used together; 10 were excluded as they were not RCTs or observational studies. 8 studies were then entered into the systematic review. Data was extracted as follows: Author and year of publication, country in which the research took place, and the sample size, study design, conditions and diseases treated with CST in the studies, treatment follow up period, the PROs used and whether the PRO was primary or secondary.

Each outcome measure was then examined to assess the extent to which it met the criteria for a measure of CST. Measures were described if they met a least one of the criteria outlined on pg. 45.

Review of studies found in search 2

Table 7 shows the data extracted on these eight studies. All studies had been conducted since 2008. The conditions studied were dementia (Gardner *et al.*, 2008); multiple sclerosis (Raviv *et al.*, 2009); fibromyalgia (Mataran-Penarrocha *et al.*, 2011); migraine (Mann *et al.*, 2012) and (Arnadottir & Sigurdardottir, 2013); pelvic girdle pain (Elden *et al.*, 2013); lower back pain (Castro-Sanchez *et al.*, 2016) and chronic neck pain (Haller *et al.*, 2016). There was much variation amongst studies regarding the follow-up period, treatment protocols, treatment frequency and duration. None of the studies were conducted in the UK, and the studies had between n = 9 (Gardner *et al.*, 2008) and n = 123 participants (Elden *et al.*, 2013).

Table 7: CST studies and outcome measures used

First Author (Year), Country, Sample size	Trial design	Condition studied and domains analysed	Treatment period/ follow-up	Outcome measure name	Importance of outcome
Gardner, *et al*., (2008), USA, n = 9.	Before and after observational study	Dementia and agitation	6 weeks Baseline then 12-week follow-up	1) Cohen Mansfield Agitation Inventory (Modified)	Primary
Raviv, G. *et al*., (2009), ISL, n = 28	Before and after observational study	Lower urinary tract signs in multiple sclerosis	4 weeks' baseline and 4 weeks post	2) Overactive Bladder (OAB-V8) questionnaire 3) QOL estimation 7-grade numeric scale	Secondary Secondary
Mataran-Penarrocha, *et al.,* (2011), ESP, n = 84	Experimental, double blind longitudinal clinical trial	Fibromyalgia • Pain intensity • Quality of life • Sleep • Depression • Trait and state anxiety	25 weeks (at baseline, 25 weeks, 6 months 1 year)	4) Visual Analogue Scale (VAS) 5) Medical Outcomes Study Short Form 36 (SF-36) 6) Pittsburgh sleep quality index (PSQI) 7) Beck depression inventory (BDI) 8) State Trait Anxiety Inventory (STAI)	Not stated
Mann, *et al*., (2012), USA, n = 69	Randomized controlled clinical trial comparing CST to low strength static magnets	Migraine • Headache QoL • Headache disability • Functional status and general health related quality of life	Baseline, 8 weeks, follow-up 4 weeks' post	9) Headache Impact Test (HIT-6) 10) Migraine disability assessment score (MIDAS) 11) Medical Outcomes Study Short form 36 (SF-36)	Primary Secondary Secondary
Elden*, et al.,* (2013), SWE, n = 123	Randomized multi centre single blinded study	• Pelvic Girdle Pain • Pain intensity • Disability • Quality of life • Pain intensity	Baseline, follow-up 1 week after treatment (8-weeks)	12) Visual Analogues Scale (VAS) 13) Oswestry Disability Index 14) European Quality of Life measure (EQ-5D)	Primary Secondary Secondary

Arnadottir & Sigurdardottir, (2013), Iceland, n = 20	Randomized controlled Train will cross-over design	Migraine • Headache QoL	12 weeks Baseline, 2	15) Headache Impact Test (HIT- 6)	Primary
Haller, *et a*l., (2016), GER, n = 54	Randomized sham controlled trial	Chronic neck pain • Pain intensity • Functional disability • Health related quality of life • Subjective physical wellbeing • Anxiety and depression • Stress perception • Pain acceptance • Body awareness • Global impression of improvement	Baseline, week 8 and 3 months	16) Visual Analogue Scale (VAS) 17) Neck Disability Index 18) Short Form 12 (SF-12v2) 19) Subjective Physical Wellbeing Questionnaire (FEW-16) 20) Hospital anxiety and depression scale (HADS) 21) Perceived stress questionnaire (PSQ-20) 22) Positive Life - Construction Scale 23) Scale of body connection 24) Patient global improvement Index (PGI-I)	Primary Secondary Secondary Secondary Secondary Secondary Secondary Secondary Secondary
Castro-Sanchez*, et al*., (2016), ESP, n = 64	Single blinded randomized control trial	Lower back pain • Disability • Disability • Pain intensity • Kinesiophobia	Baseline, after treatment and one-month follow-up	25) Roland Morris Disability Questionnaire 26) Oswestry Disability Index 27) 10-point numeric pain rating scale 28) Tampa Scale of Kinesiophobia	Primary Secondary Secondary Secondary

Identifying the PROs which are most relevant to CST

Twenty-five different outcome measures (listed in *Table 7*) were used as part of the study designs to assess primary and secondary outcomes. They were grouped by type as described in the 'Typology of measures' section, (pg. 41) to confirm they met the requirements to capture CST outcomes. Nine of the outcome measures were disease specific (1, 2, 7, 8, 9, 10, 13, 15, 17, 20), five measured health related quality of life (HRQoL) (3, 5, 11, 14,18) and the remainder assessed pain intensity (4, 12, 16, 18 and 27), pain acceptance (22), sleep quality (6), perceived stress (21), body awareness (23) subjective physical wellbeing (19), impression of improvement (24), Kinesiophobia (28) and disability (13, 17, 25 and 26).

Having applied the criteria shown on pg. 45 each questionnaire was appraised to see if they covered more than one domain. The disease/condition specific measures and those measuring symptoms used in the studies under review have not been described here as these instruments focus on a single domain. Of the remaining instruments, the QoL of estimation scale (3) was a one-item scale and did not provide sufficient provision for CST. Three of the HRQoL questionnaires met at least some of the criteria laid out on pg. 38 (SF-36, SF-12 and EQ-5D) and these are described in greater detail.

Health-related quality of life measures used in CST studies to date

Both the SF-36 and SF-12v2 fail to capture outcomes of a spiritual nature, and do not address effects linked to self-concept in the mental wellbeing domain. Although the main emphasis is on functional and emotional status, the SF-36 and SF-12v2 lack items that show how responders have or have not adapted to their situation and to address their ability to cope as mentioned on pg. 23. Neither have items which tap self-awareness or self-care. Although lacking anything about spiritual wellbeing they are the most holistic of the measures identified. Both are well validated and well respected in conventional medicine and because of this, both are looked at

in more detail shortly. The EQ-5D has items relating to mental and physical health, and one item specific to self-care, for this reason it was selected for review. However, the EQ-5D does not capture spiritual or social outcomes and does not address mental wellbeing.

36 Item Short Form Survey (SF-36)

The Short Form Survey (SF-36) (Ware, 1992) is a 36 -item multi-dimensional questionnaire that assesses 8 dimensions of health status covering physical mental and social health. The domains are labelled: physical functioning; role limitations due to physical health problems; bodily pain; social functioning; general mental health; role limitations due to emotional problems; vitality, energy or fatigue; and general health perceptions. It has a recall period of four weeks and can be completed by self or with an interviewer. The items were derived from parent questionnaires SF-18 and SF-20 created by clinicians (Ware & Sherbourne, 1992) for the US Medical Outcome Study to survey health status. Scores represent the total percentage possible score achieved. It has been developed for adults and is available in multiple languages. Validation of the SF-36 in the UK was carried out by Brazier *et al*., (1992) in a general practice setting. Table 8 shows the internal consistency and test-retest reliability results, the reliability of the SF-36 is good as Cronbach's $\alpha > 0.85$ for all dimensions except for social functioning and test-retest reliability results are good. With > 90% of cases being within 95% confidence interval in all domains. Construct validity was evaluated using the Nottingham Health Profile (Kind & Carr-Hill, 1987).

Table 8: SF-36 internal consistency and test-retest reliability

		Internal consistency		Test-retest reliability (2 weeks)	
Domain/Scale	**Items**	**Alpha**	**Correlation**	**Mean difference**	**% of cases lying within 95% CI**
Physical functioning	10	0.93	0.81	0.49	98
Role functioning/physical	4	0.96	0.69	0.57	98
Role functioning/emotional	3	0.96	0.63	0.44	97
Vitality	4	0.96	0.80	0.39	96
Mental health	5	0.95	0.75	0.71	91
Social functioning	2	0.73	0.60	0.15	93
Pain	2	0.85	0.78	0.71	95
General health	5	0.95	0.80	0.40	96

Adapted from (Brazier *et al.*, 1992).

The SF-36 is used internationally and has been applied in many healthcare settings (Gandek *et al.*, 2004), including use in CAM e.g. (Braun *et al.*, 2011; Cooley *et al.*, 2009; Lee *et al.*, 2013; Skillgate *et al.*, 2010; Wong *et al.*, 2011).

12 Item Short Form Health Survey (SF-12)

The SF-12 health survey is a subset of the SF-36. It is a 12-item instrument designed to measure overall mental and physical health. Items that assess mental health, role emotional and social functioning and vitality go to make up the mental component summary (MCS) which yields a single score of overall mental health. Items associated with physical functioning, role physical, bodily pain and general health go to make up the physical component summary (PCS) and generate a single score of overall physical health. The instrument was designed for use in the adult population, and is self-completed or with an administrator. Two versions are available, the first has a recall period of 4 weeks and the second is for use in acute situations and has a recall period of 1 week.

In studies carried out in the UK, the SF12 has internal consistency reliability estimates of 0.91 for the PCS measure and 0.87 for the MCS measure (Ware *et al.*, 2010). And have a test-retest for the PCS, 0.86 and MCS, 0.77

respectively (Ware *et al.*, 1995). Items are added in each scale to obtain a raw score. The raw score is then transformed to a 0-100 score; this is then transformed to a *z* score, and the scale *z* score is then transformed to a *T* score (mean=50, standard deviation =10). Computer-based scoring services are available through QualityMetric Incorporated. The SF-12 was updated in 2002 (Ware *et al.*, 2002) (SF-12v2). Refinements to wording and layout were made to make it easier to complete. The SF-12 has been used in studies of CAM (Eisenberg *et al.*, 2012; McKee *et al.*, 2012).

EQ-5D-5L

The EQ-5D-5L is a two-part instrument which includes the EQ-VAS (Euroqol Group, 1990; Herdman *et al.*, 2011). Part one is a 5-item scale that is used as a generic measure of health outcomes applicable to many health conditions and settings. The descriptive scale addresses mobility, self-care, usual activity, pain/discomfort and anxiety/depression. Response options include: no problems, slight problems, moderate problems, severe problems, and extreme problems. Responders assign a level of severity (1-5). These numbers are combined to create a five-digit number which describes the respondent's health (Euroqol Group, 2015) as a utility value; a health economic evaluation given to different health states (defined so that death has a value of zero and perfect health is given the value of 1 (Euroqol Group, 2015). Part two contains the VAS health thermometer which records an individual's assessment of their own health status. This questionnaire is for adults and designed for self-completion. No recall period is required; the answers reflect the state of the individual at time of completion.

The initial descriptive system of the EQ-5D was not developed using modern psychometrics (Devlin & Brooks, 2017) although face and content validity of the EQ-5D-5L version has been evaluated using focus groups in the UK (Herdman *et al.*, 2011) with the aim of reducing ceiling effects. A ceiling effect occurs when a questionnaire has a clear upper limit for

potential responses and many participants score at or near this limit (Lewis-Beck *et al.*, 2004). Test-retest reliability was analysed using inter-class coefficients, ICC: 0.78 at individual level and at group level using the Wilcoxon Tests: 0.73 (Brooks, 1996). The EQ-5D has been used in the evaluation of CAM (Thomas *et al.*, 2005; Weze *et al.*, 2004).

2.4.3 Search 2 summary

Eight studies were reviewed to identify PROs applied in CST studies; twenty-five different outcome measures were used within these studies to assess primary and secondary outcomes. Of these, only three HRQoL questionnaires (SF-36, SF-12 and the EQ-5D) met with some of the criteria for a CST PRO and were reviewed.

None of the instruments identified in search 2 matched all of the areas of outcome important to CST users (none had provision for spiritual wellbeing, or items that evaluate self-awareness). In the light of the findings of the first two reviews a further search was undertaken focusing on PROs developed for use in CAM.

2.5 Search 3: identifying and appraising PROs developed for CAM

The aim of this search was to identify and appraise patient-reported outcomes developed for use in CAM.

2.5.1 Methods for search 3

A systematic search was undertaken in 2014 and updated in December 2016. Electronic databases (Medline, AMED and EMBASE databases *via* OVIDSP, CINAHL *via* EBSCOhost, PsycINFO and PsychTests *via* ProQuest and ScienceDirect) were used to identify published papers. Reference lists were checked to ensure articles of relevance known to researchers in the field were not missed.

Table 9: Search 3, search terms, strategies and results (2016 search)

Search	Search terms	Results
1	exp "Outcome Assessment (Health Care)"/	906546
2	exp "Surveys and Questionnaires"/	884480
3	database*.mp.	322813
4	1 and 2 and 3	3455
5	(patient reported outcome measure* or primary care outcome measure* or patient reported outcome questionnaire*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	1617
6	2 or 5	885093
7	3 and 6	21797
8	(questionnaire* or database*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]	870354
9	measurement properties.mp.	1652
10	terwee*.au.	254
11	9 and 10	49
12	filter.mp.	54002
13	11 and 12	1
14	alternative therapy.mp. or exp Complementary Therapies/	222391
15	7 and 14	246

The following inclusion and exclusion criteria were applied to the list of articles:

Inclusion criteria

i. Papers describing the development of a PRO for CAM
ii. Systematic reviews of PROs suitable for CAM

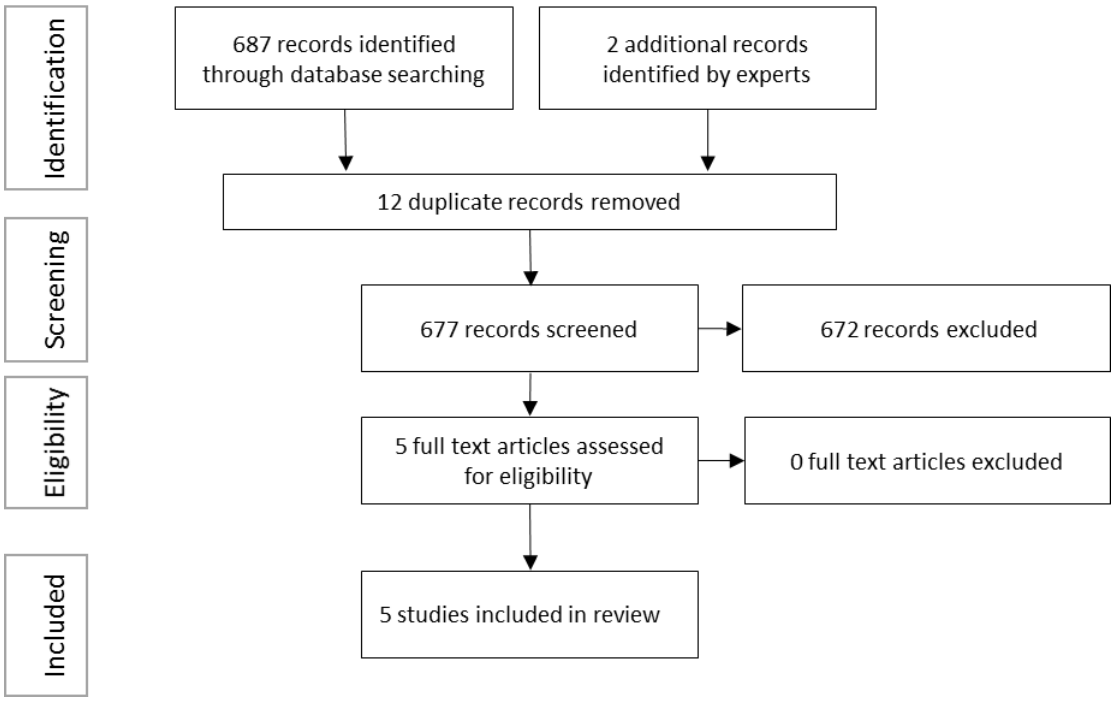
Exclusion Criteria

i. Not published in English language
ii. Not published in a peer reviewed journal
iii. Measures developed or evaluated for patients > 18 years

2.5.2 Results of search 3

Figure 5 shows the number of records identified, screene and retrieved to assess for eligibility and those studies included in the review.

Figure 5: Flow of information through the systematic search of PROs for CAM



The search resulted in 689 records that referred to PRO use in CAM. Twelve duplicates were removed and 677 abstracts were screened. Six hundred and seventy two records were excluded as they did not meet the inclusion criteria, the number of eligible full text articles totalled 5. Of the five studies identified, four were linked to the development and validation of two patient-reported outcome measures for CAM: the Harry Edwards Healing Impact Questionnaire (HEHIQ) (Bishop *et al.*, 2010) and the Self-Assessment of Change (SAC) (Ritenbaugh *et al.*, 2011). One study was a systematic review of patient-reported outcome measures for integrated medicine practices in primary care (Hunter & Leeder, 2013). In the first search in 2014 only the HEHIQ and the systematic review were identified. The SAC was only identified in the 2016 search.

Systematic review of patient reported outcome measures for integrated medicine practices in primary care

The systematic review (Hunter & Leeder, 2013) set out to identify, appraise and short-list the best available patient-reported outcome measures for integrated medicine practices in primary care within Australia. Two search strategies were undertaken: the first was to identify web-based databases which list PROs potentially suitable for CAM and the second search aimed to identify aspects of health and wellbeing that were under-represented on the PRO databases.

This review identified 10 web-based databases as listing PRO questionnaires. All questionnaires listed on these databases were combined yielding approximately 3800 instruments. The following exclusion criteria were then applied: not patient completed, not paper/electronic format, non-English only, not for all adults ≥ 18 years of age, disease, symptom or treatment, specific, only for use in hospitals or Institutions, not measuring patient attributes. This resulted in 478 questionnaires. The second search strategy identified 168 PRO questionnaires that had under-represented items related to health and wellbeing, such as smoking cessation and nutrition. Of these 646 questionnaires, 575 were excluded and 71 were appraised using the Medical Outcomes Trust appraisal framework (Lohr, 2002).

Hunter's (2013) systematic review informed the searches used in this thesis in two ways: 1) an attempt was made to identify current databases of PROs to update the search for PROs developed for use in CAM, and 2) the list of PROs recorded in the review as suitable for use in studies of CAM was scrutinised to identify measures developed specifically for CAM.

Identifying current databases of PROs

On September 15th 2014, a replication of Hunter's (2013) first search strategy was attempted by searching the internet for PRO databases. Using the search engine Google, Table 10 shows the search terms used:

Table 10: Internet search terms

Search terms	"patient reported outcome", "patient reported outcome database", patient reported outcome database", "outcome scales", "health outcome database".

The first ten pages of the Google search results were checked. N=7 databases were identified along with multiple sites advertising specific suites of outcome measures, (e.g. www.euroqol.org) (EuroQol Group, 1990) which maintains the EQ-5D (Euroqol Group, 1990) suite of outcome measures) all of which were identified by Hunter (2013).

Of the ten databases identified by Hunter (2013), the websites of the Outcomes Scale Repository, Chartered Society of Physiotherapy and PRO Measures are no longer available. Two were out of date: the Medical Outcomes Trust (MOT., 1992) was last updated in 2006 and the PROM Group (NCHOD., 1998) was last updated in 2005. Four databases remained active: PROQOLID (Mapi., 2002), Australian Centre on Quality of Life (ACQoL) (Deakin University, 2002), World Happiness Database (Veenhoven, 2003) and IN-CAM outcomes database (Verhoef *et al.*, 2007). It was not possible to identify when new measures have been added to the PROQOLID, ACQoL and World Happiness Databases, making updating searches impossible. The Patient-Reported Objective Measurement Information System (PROMIS) database (Healthmeasures, 2004) is an item bank of questions and is, therefore,

not a suitable source in this context. The IN-CAM outcomes database described on pg. 37 was the only source suited to this study and identified as being actively updated.

The IN-CAM database was scrutinised during March and April 2014, and again in March 2017 to identify potential candidate questionnaires for CST. A total of 279 questionnaires were listed on the database in March 2014 and 276 in March 2017. Development and validation papers were sourced and read for questionnaires, which included the domains of importance to CST users (Brough *et al.*, 2015): social, spiritual, physical and psychological.

As mentioned in Section one, questionnaires on the IN-CAM database were categorized by nine domains: 1) health related quality of life, 2) individualized, 3) holistic, 4) social, 5) spiritual, 6) physical, 7) psychological, 8) process of the intervention and 9) context of the intervention. Many of the questionnaires were disease specific, such as the Wellbeing Questionnaire-28 (WBQ-28) (Speight *et al.*, 2000) developed for use in diabetes. Although, in areas of health care where CAM are commonly used such as oncology and chronic illness (FACIT) (Winstead-Fry & Schultz, 1997), references which included CAM were also available (Hoffman *et al.*, 2012). Whilst a full review of the IN-CAM database is beyond the scope of this thesis, it demonstrates that PROs within the field of CAM are limited and strengthens the case for a novel therapy-specific questionnaire for CST.

2.5.3 A) PROs developed for CAM from the Hunter review

The 71 PRO questionnaires included in the Hunter (2013) review were reviewed to identify measures developed for CAM.

Two questionnaires developed specifically for CAM were identified: Complementary and Integrative Medicine Outcomes Scales (CIMOS) (Eton *et al.*, 2005) and the Measure Yourself Medical Outcome Profile (MYMOP) (Paterson, 1996a).

Complementary and Integrative Medicine Outcome Scales (CIMOS)

The CIMOS (Eton *et al.*, 2005) has 29 (35) items and is made up of seven domains: pain (5 items), fatigue (5 items), physical/functioning ability (5 items), personal control (6 items), existential (6 items), general quality of life (2 items) and practitioner skill (6 items). The intention is that the seven CIMOS domains can be included or excluded depending on the aims and objectives of a research study. Responses are captured on a five-point Likert scale (Likert, 1952) ranging from none of the time (0), to all the time (4), for the domains of pain, fatigue and physical functioning ability. The remainder of the domains; personal control, existential, general quality of life and practitioner skills are rated on a five-point Likert scale (Likert, 1952) ranging from zero (strongly disagree), to four (strongly agree). The CIMOS has a recall period of one week. Scores are calculated by summing the items in each domain. There is no 'total score', as the CIMOS has been developed to allow researchers to choose domains that match their study aims and objectives, thus making comparisons across studies challenging.

Validation of the CIMOS

The CIMOS was developed by pooling items from five well validated instruments to create a draft questionnaire. Qualitative methods were undertaken with both practitioners and users of CAM (n = 42). In addition, researchers and an outcome measurement specialist were involved in evaluating each item for inclusion on the pilot questionnaire to ensure good content and face validity. A quantitative validation was carried out with a population of n = 52. Only six of the seven domains were evaluated, with the practitioner skills domain not reporte. Eton *et al*., (2007) report that the internal consistency reliability, Cronbach's α, for the domains as: Pain α = 0.96, fatigue α = 0.80, physical/functional α = 0.81, personal control α = 0.72, existential α = 0.79, general QOL α = 0.68. Multi-trait scaling analysis supported some item convergence and item discrimination in the six domains. The CIMOS was checked against external standards to evaluate construct validity, using domains from the Brief Pain Inventory (BPI) (Cleeland, 1989), the Brief Fatigue Inventory (BFI) (Mendoza *et al.*, 1999), an adaptation of the Selby's Linear analogue self-assessment uniscale (Moinpour, 1994) and the Positive and Negative Affect Scale (PANAS) (Watson *et al.*, 1988).

Psychometric analyses (by either factor analysis or item response theory) were not carried out. Test-retest was not evaluated and responsiveness was not assessed.

Whilst conceptually the CIMOS appears to cover many of the outcomes reported as important to CST users, it fails to match personal control as a mental wellbeing component in the domain of mental health. There is, however, much emphasis on pain and fatigue with 10 of the 29 items focusing on these topics in the CIMOS, whilst symptoms (pain) only represented a small part of the overall content of the conceptual framework of CST outcomes. In addition, the CIMOS has two differing sets of response options on the questionnaire and this may be problematic and confusing to responders. The process of evaluating the measurement properties of the CIMOS was limited due to the small sample size and lack of diversity within the sample. The alpha value of 0.96 for the pain domain is outside of the acceptable range (>0.70 to 0.90) and may indicate that there are redundant items within that domain and an increase of patient burden. The lack of psychometric analysis and tests of responsiveness mean that the structure, reliability and validity of the CIMOS are yet to be established. To date, there are no published studies using the CIMOS. Thus, the CIMOS was judged as unsuitable at present to evaluate CST outcomes.

2.5.4 Measure Your Medical Outcome Profile (MYMOP)

The second PRO developed for CAM use identified in the Hunter review (2013) was the Measure Your Medical Outcome Profile (Paterson, 1996a). The MYMOP is a patient-centred measure with six items at a baseline and seven items at follow-up. It caters for two patient-selected symptoms either physical or mental, and asks four questions for which responses are marked on a 0 - 6 Likert scale (Likert, 1952) (0 being good and 6 being bad). Patients are asked to choose one activity (physical, social or mental) to score how bad it has been and they are also asked to rate their 'general feeling of wellbeing'. A scale is produced for each problem. Responders are asked about their medication related to their chosen symptoms. The MYMOP is designed to be administered by the practitioner and the recall period is one week, however, the MYMOP2 has also been validated to be self-administered (Paterson, 2000). The profile score is calculated as the mean of the scored items.

Validation of MYMOP

The MYMOP has been used successfully to evaluate patient outcomes for both allopathic (n = 218) and CAM (n = 47) (Paterson, 1996b). Traditional psychometric methods have not been used, but construct validity comparisons have made with the SF-36 health survey and responsiveness has been evaluated. The MYMOP was used in a general practice consultation with practitioners and repeated after four weeks. A consistent grade across the spectrum of clinical change between scores from 'a little better' to 'about the same' were shown as significant on all the MYMOP scales except 'wellbeing' suggesting that the MYMOP is responsive. Qualitative studies to explore treatment effects of patients using the MYMOP are available (Paterson, 2004). The MYMOP has been used in numerous CAM studies (Cooley *et al.*, 2009), (Chapman *et al.*, 2001; Jackson *et al.*, 2006), (Abbot *et al.*, 2001; Cleland *et al.*, 2006; Paterson *et al.*, 2005; Thompson *et al.*, 2005) including a study of CST published in a non-peer reviewed journal (Isbell & Carroll, 2007).

A measures design where outcomes that individuals report as being important to them, is in keeping with the CAM ethos; yet use of the MYMOP in a homogenous population with the same condition creates challenges. The MYMOP is unable to measure change relating to emerging problems or in several co-existing problems and concerns; failing to capture unexpected change in new symptoms or symptoms beyond the two identified by the responder as important. This may be too narrow a focus to the overall situation of the patient despite being a patient-centred approach. In the context of CST outcomes, the MYMOP could capture symptoms of mind and body, yet does not address the spiritual outcomes reported as meaningful to clients (Brough, 2012).

2.5.5 B) Other PROs developed for CAM identified from systematic search

The remaining papers (Bishop *et al.*, 2010; Paterson *et al.*, 2007; Ritenbaugh *et al.*, 2011; Thompson *et al.*, 2011) identified in search 3 but not featuring in the Hunter review, identify the validation of the Harry Edwards Healing Impact Questionnaire, the Measure Your Concerns and Welbeing (MYCaW) and the Self-Assessment of Change (SAC) questionnaire.

Harry Edwards Healing Impact Questionnaire (HEHIQ)

The Harry Edwards Healing Impact Questionnaire (Bishop *et al.*, 2010) is a therapy-specific measure designed for spiritual healing. The questionnaire was developed using a mixed methods approach. Both users and practitioners of spiritual healing had input into the development of the HEIHIQ. Focus groups were used to identify effects of spiritual healing, from which draft items were generated. Focus groups and cognitive interviews helped to refine the items. The HEHIQ has both 29-item (long subscales) and 20-item (short subscales) forms. Scores on the short form range between minimum 20 and maximum 100, with a high score representing a more positive state of health. Both forms assess outcomes in five domains: outlook, energy, health, relationships and emotional balance.

Psychometric overview of the short form: Internal consistency; Outlook (4 items) $\alpha = 0.83$, Energy (3 items) $\alpha = 0.85$, Health (4 items) $\alpha = 0.82$, Relationships (3 items) $\alpha = 0.73$, Emotional Balance (5 items) $\alpha = 0.89$, all items (20) $\alpha = 0.91$. Test-retest reliability: over 2 weeks, showed poor reliability (all items mean = 0.00, SD = 0.29). Sensitivity to change: change was seen on all scales except outlook and relationships, all items were reported to have mean magnitude of difference of 0.39 and SD of difference 0.36 (Bishop *et al.*, 2010). The HEHIQ has yet to be tested in a homogenous population with the same condition. To date, there are no studies published beyond the development and validation papers.

The HEHIQ has been validated in a population like that of CST, making it a useful comparator measure and Brough (2012) hypothesised that the domains included in the HEHIQ would be applicable in CST, yet lacked items referring to the physical body in the way that CST users had described.

Measure Yourself Concerns and Wellbeing (MYCaW)

An individualised PRO designed for measuring the effects of cancer support services which include complementary therapies is the Measure Yourself Concerns and Wellbeing (MYCaW) (Paterson *et al*., 2007). This questionnaire is an adaptation from MYMOP (Paterson, 1996) see pg.65. Having the same format, two questions are scored on seven-point scales (with higher scores representing worse wellbeing). As described for the

MYMOP, instead of using the term 'symptoms' (either 'physical or mental') the terms 'concerns or problems' have been adopted without the need to specify what the concerns or problems are related to. Individuals are also asked to rate their general feeling of wellbeing. They are also asked for qualitative data about 'other things that may be affecting your health' and 'the most important aspects of the therapy'. The MYCaW does not include a score for a question about activity and a further modification to MYMOP on the follow-up form asks the open question 'Reflecting on your time with this Centre, what were the most important aspects for you?' The MYCaW can be self-completed or used face to face with an administrator. The profile score is calculated as the mean of the scored items and the absolute difference between the baseline and follow-up scores describes the amount of improvement or deterioration. Qualitative data were analysed and coding guidelines developed (Polley *et al*., 2007) to aid researchers with qualitative analysis and to facilitate comparisons across other published data sets. The content of these guidelines will be discussed later in the thesis, see pg. 215.

The MYCAW does not present a vision of the areas of health or wellbeing that might change as a result of therapy. Its strength is, therefore, that it can capture change that researchers had not envisaged. On the other hand, it also relies on participants' knowledge and awareness of what might change to capture the full extent of change.

2.5.6 Self-Assessment of Change (SAC)

Self-Assessment of Change (Ritenbaugh *et al.*, 2011) was developed using a mixed methods approach following FDA guidance (2009). The SAC has six domains that consist of 18 pairs of terms, anchored at both positive and negative ends of each domain. The domains include physical, cognitive, emotional, social, spiritual and whole person characteristics. A 100-mm visual analogue scale (VAS) connects the items. Responders mark on the VAS with a 'B' where they were before the intervention and with an 'N' to indicate where they are currently. The lines are measured from the left edge to B and N for data entry. The SAC can be completed by paper or online.

The SAC was evaluated in 7 studies in CAM settings using both paper and online formats, n = 519, the scoring method is reported as problematic if using a paper version (Ritenbaugh, 2014). There is no coherent quantitative component to this evaluation nor, currently, any peer review paper to evidence its evaluation. A retrospective pre-test approach (Howard *et al*., 1979) was used to evaluate the data to overcome response shift in this population. Test-retest reliability has yet to be evaluated.

Whilst the SAC captures some outcomes relevant to CST, it does not capture outcomes related to self-care or patient engagement.

2.6 Consultation with experts

As a final step in the search for PROs, academics active in research in CAM were consulted and an edition of the European Journal of Integrative Medicine dedicated to outcome measurement was scrutinised. Three measures were recommended in this process: the SAC, the HEHIQ and the Warwick-Edinburgh Mental Wellbeing Scale (Stewart-Brown *et al.*, 2015). The first two of these measures have already been covered because they came up in the search updated in 2017. The third was not specifically developed for use in CAM but a strong case had been made that it was likely to be useful in this context and it covered a domain of importance for CST which was not covered by other measures.

2.6.1 Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS)

The Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) is a measure of positive mental wellbeing (Tennant *et al.*, 2007) comprising 14 positively worded statements. The scores range from 1 – 5 on a Likert scale (Likert, 1952) with response categories ranging from 'none of the time' to 'all of the time'. The minimum total score is 14 and the maximum is 70, with higher scores representing higher levels of mental wellbeing. The recall period is 2 weeks and may be self-administered.

The WEMWBS was initially validated in a population based on student samples and subsequently in focus groups held in Scotland and England (Tennant, *et al.*, 2006) and among people aged 13 to 74 in the UK (Clarke *et al.*, 2011; Tennant *et al.*, 2006). Internal

consistency: α = 0.89 (n = 348) suggests that there is a good level of internal consistency. Test-retest reliability: Intra-class correlation coefficient (ICC) = 0.83 after one week (n = 124). The WEMWBS was checked against external standards to evaluate convergent validity using the Scale of Psychological Wellbeing, Satisfaction with Life Scale, Short Depression Happiness Scale, Positive and Negative Affect Scale (positive subscale), WHO-Five Wellbeing Index, the EQ-5D thermometer and the Emotional Intelligence Scale. Sensitivity to change: the WEMWBS has been shown to be responsive at both group and individual levels (Maheswaran *et al.*, 2012). The WEMWBS has been used in studies evaluating CAM (Hagins, 2016) and (Korpela, 2016).

When appraising the contents of WEMWBS against the outcomes of importance, topics related to 'being able to connect with others' and 'coping better' matched those reported in previous qualitative work (Brough *et al*, 2012). Despite being a measure developed for positive mental wellbeing and psychological functioning, one aspect, 'I've been feeling closer to other people' may overlap into other concepts as CST users reported 'feeling connected in relationships' and 'feeling connected to the wider universe' (Brough, 2012, pg. 117) which were categorised as spiritual outcomes, thus highlighting a grey area and the challenge of defining themes. The WEMWBS does not capture the physical outcomes and would need to be combined with a measure of physical health if used in a population of CST users.

2.7 A synopsis of the eight questionnaires reviewed in this thesis

Eight questionnaires were reviewed in this thesis, four developed for CAM, the CIMOS, HEHIQ, SAC questionnaire and the MYMOP and four developed in non CST/CAM settings, the WEMWBS, SF-36, SF-12v2 and the EQ-5D.

Tables 11 and 12 lists each of these questionnaires and shows the domains of importance required for CST outcomes and which questionnaires tap these domains. It presents the number of items and provides a rating (good, moderate or poor) for the reliability, validity and sensitivity to change of each questionnaire. A comments section focuses on the areas

in which the questionnaires lack provision or aspects which impact suitability for CST and any applicable costs are reported.

Summarising the content of Table 11 and Table 12: whilst generally there was some matching with the content of measures developed for CAM use, each measure lacked provision to capture all the outcomes of relevance to CST in some way. The CIMOS has a high emphasis on pain and fatigue items; the HEHIQ lacks items related to physical health, self-awareness and self-care but the remainder items tap the other outcomes of CST; the SAC lacks items that tap self-awareness and self-care; the MYMOP lacks provision for outcomes of a spiritual nature. The reliability and validity of these questionnaires were questionable as only the authors have tested the questionnaires and the results suggest either poor reliability and validity or that the psychometric properties are yet to be established. The MYMOP is the only questionnaire in this group which has demonstrated the ability to detect real change over time.

Referring to the content of non-CST/CAM questionnaires: the WEMWBS captured positive mental wellbeing only and lacks items which address the physical functioning, self-care and self-awareness; SF-36 and the SF-12 v2 lack provision for spiritual outcomes; they do not address adaptation or coping, or have items that tap self-awareness. These measures did not incorporate patient input during the early stages of their development and have a high emphasis on symptomatology; the EQ-5D also lacks provision for spiritual outcomes and self-awareness. All four questionnaires developed in non-CST/CAM settings have good reliability and validity and are sensitive to change. Fees are applicable for the SF-36, SF-12 and EQ-5D depending on application; no fee applies to the WEMWBS.

To conclude: all questionnaires assessed in this review lacked provision in at least one area, and failed to tap items for self-awareness and most failed to tap self-care. The development of a new PRO to capture all outcomes of importance to CST users is, therefore, justified.

Table 11: Candidate PROs (non-CAM)

PROs non-CAM	Domains of importance: Physical WB, Mental WB, Spiritual WB, Social WB, Self- awareness, self-care, patient engagement	No of items	Reliability [a]	Validity [a]	Sensitivity to change [b]	comments
Warwick - Edinburgh Mental Wellbeing Scale (WEMWBS)	Positive affect, relationships and functioning	14	Good	Good	Yes	Lacks domains of physical, items of self-care and self-awareness, responsibility for self. Free to use. Total Score.
Short Form Health Survey 36 (SF-36)	General, physical, pain, energy, mental, social, and role functioning.	36	Good	Good	Yes	Lacks spiritual domain. Fails to address adaptation and coping, patient engagement. No patient input during development. Cost for use.
Short Form Health Survey 12 (SF-12)	Physical functioning, Role-physical, bodily pain, general health, vitality, social functioning, role emotional, mental health	12	Good	Good	Yes	Lacks spiritual domain. Fails to have address adaptation and coping, patient engagement. No patient input during development. Cost for use.
Europe Quality of Life Scale (EQ-5D)	Pain, activities, mobility, self-care, VAS overall health and utility score	5 and 1 VAS	Good	Good	Yes	Lacks spiritual domain, self-awareness. Large ceiling effects and some floor effects. Cost for use.

[a] A rating of 'good' means the questionnaire has been tested in many studies conducted by researchers other than the author, with measurement properties reported as 'acceptable' confirming the different reliability and validity. A rating of 'poor' means that only the authors have tested the questionnaire and the results suggest poor reliability or validity, or they have not been tested. 'Moderate' means the quality of the evidence is somewhere between 'good' and 'poor'.

[b] Questionnaires had demonstrated the ability to detect real change over time.

Table 12: Candidate PROs developed for CAM assessed in this thesis

PROs developed for CAM use	Domains of importance: Physical WB, Mental WB, Spiritual WB, Social WB, Self-awareness, self-care, patient engagement	No of items	Reliability [a]	Validity [a]	Sensitivity to change [b]	comments
Complementary and Integrative Medicine Outcome Scale (CIMOS)	Pain, fatigue, physical, personal control, existential issues, QoL.	(29) 35	Poor	Poor	Unknown	High emphasis on pain and fatigue items. Free to use. No total score given. Modular domains
Measure Yourself Medical Outcomes Profile (MYMOP)	Identifies problem, rates two symptoms and one activity and change. Also, records change in medication use.	7	Moderate	Good	Yes	Lacks spiritual domain. Free to use, Measures two problems/illness only.
Harry Edwards Healing Impact Questionnaire (HEHIQ)	Mental outlook, energy, health, relationships, emotional balance	20	Poor	Poor	Unknown	Was developed and evaluated in a sample similar to that of CST. Lacks items related to physical health, self-awareness, self-care and responsibility for self. Free to use. No total score given.
Self-Assessment of Change (SAC)	Sleep, Energy, Senses, Focus, overwhelm, hope, connection, mood, defined.	18-word pairs	Moderate	Poor	No	Lacks items on self- awareness, self-care and responsibility for self. Nine-word pairs, responders rate change between now and before on a Visual analogue scale.

A rating of 'good' means the questionnaire has been tested in many studies conducted by researchers other than the author, with measurement properties reported as 'acceptable' confirming the different reliability and validity. A rating of 'poor' means that only the authors have tested the questionnaire and the results suggest poor reliability or validity, or they have not been tested. 'Moderate' means the quality of the evidence is somewhere between 'good' and 'poor'. [b] Questionnaires had demonstrated the ability to detect real change over time.

2.8 Methodological issues of searching for PROs for CST and CAM

Searching for PROs is different to searching for RCTs or other research on CST and CAM. Systematic searches for CAM PROs are possible, yet fail to identify all the relevant literature (e.g. omission of the SAC during the search in 2014). Various factors contributed to the limitations of searches in this area. A replication of search strategy 1 from Hunters review (2013) was unworkable since it was not possible to identify new measures on the PRO web databases and because these PRO databases a) cannot be searched by the date when measures were added; b) are often not kept up to date and c) are not always available. It was not possible to identify any new PRO measures within a specific search time frame. Manual counts of database content proved to be laborious, time consuming and yielded unclear results.

The search strategy used in search 3 when identifying PROs developed for CAM resulted in tens of thousands of papers being listed on the electronic medical databases with a clear majority being irrelevant to this study. Hence, keyword searches in PRO development are unreliable as there does not appear to be consistency in the way in which articles are indexed. As this is dependent on the indexers and the possibility of human error occurring, some secondary or minor subjects may not be indexed and the level of indexing can vary amongst databases. During the 2017 search update, a PubMed filter for finding measurement properties of measurement instruments was found (Terwee *et al.*, 2009) and used in the updated search 2017. Had this filter been identified in 2014, the search process might have been less challenging.

2.9 Chapter summary

A series of searches were undertaken to discover if the planned research was necessary or appropriate. These were designed pragmatically to optimise the chances of finding existing measures that might be

appropriate within the limited time frame available for this thesis. The first search aimed to identify PROs developed for CST and the second to identify PROs applied in CST studies. A third search was undertaken to identify PROs developed for CAM. The results revealed that 1) no PRO had been developed specifically for CST; 2) three questionnaires used in CST studies were found and appraised; 3) a systematic review of PROs in integrated care highlighted three candidate measures for appraisal; and 4) two further PROs were identified during the search. A description of each candidate PRO and their measurement properties were reported. Each PRO was assessed to see if their domains captured the outcomes of interest for the CST population. Based on these results, the decision to develop and evaluate a CST specific questionnaire was made. The section concluded with a section on the methodological issues encountered when attempting systematic reviews to identify PROs developed for CST and CAM.

3 Methodology

3.1 Chapter Overview

This section considers ethical issues, bias and the process of reflexivity. The standards and guidelines which have been adhered to in order to uphold rigour in this thesis are presented, in addition to the 'gold standard' FDA guidance (2009) on PRO development which has been adopted for the design of this study. Qualitative and quantitative methodologies are introduced and the rationale behind the choice of methods for this mixed methods research process are described.

3.1.1 Ethical issues

Ethical tensions arose in the context of research being carried out by a practitioner of CST. The researcher (NB) had to consider her conflicting roles and the responsibilities she had to the different stakeholders relating to this study. The stakeholders or multiple communities of practice (Costley & Gibbs, 2006) in this instance: CST users, CST practitioner peers/colleagues, the University and the funders, namely the CSTA and the University, had their own agendas and the researcher had to navigate these relationships during the course of her study.

3.1.2 Bias

The researcher's role as a CST practitioner introduced potential bias to this study; yet allowed her to apply insight, helped in the facilitation of focus groups and interviews and in the interpretation of the participants' sharing activities. Krueger (1994) calls this activity playing the role of the "enlightened novice". Subject bias due to the recruitment of clients *via* practitioners could potentially be problematic as clients may want to please their practitioner by taking part in the qualitative research processes and or by completing the questionnaires in a positive way. Bias

can arise in all aspects of the research process and research decisions have been made with this in mind.

3.1.3 Reflexivity

According to Banister *et al*., (1994, pg. 149) *"reflexivity is the most distinctive feature of qualitative research"*. Reflexivity is the process of critically reflecting on the kind of knowledge created and how that knowledge is generated in research (Guillemin & Gillam, 2004). This level of scrutiny can improve the quality of the research by being able to step back and view one's position within the research, being aware of possible influences and any limitations the study outcomes may have.

3.1.4 Sampling

Random samples are defined by Clark-Carter as *"those in which each population element has an equal probability, or a quantifiable probability of being selected"* (1997, pg. 165). Random sampling is not applied in qualitative methods, yet the participants are required to reflect the population of interest (Stewart & Shamdasani, 1990). Purposive sampling (Glaser & Strauss, 1967) occurs when participants are selected according to the needs of the study; and those who do not meet the profile are rejected. As in the case of Nassar-McMillan and Borders (2002), this study considers the intergroup heterogeneity and intragroup homogeneity of focus group participants to enable comparisons to be made during analysis. Purposive sampling was used in this study.

3.1.5 PRO development

There are no set international guidelines on how to develop a PRO. However, the US Food and Drug Agency (2009) has made recommendations which are considered the 'gold-standard' approach to developing and evaluating a new PRO, which combines both quantitative

and qualitative methods (Gorecki *et al.*, 2013; Ritenbaugh *et al.*, 2011). These methods were adopted.

3.1.6 Reporting standards and guidelines

Reporting guidelines for studies which develop outcome measures include: Standards for Educational and Psychological Testing (American Educational Research Association, 1999), the Standards for Reporting Diagnostic Accuracy (STARD) initiative (Bossuyt *et al.*, 2003), the Guidelines for Reporting Reliability and Agreement Studies (GRRAS), (Kottner *et al.*, 2011) and the COnsensus-based Standards for the selection of health Measurement INstruments known as the COSMIN checklist (Mokkink *et al.*, 2010). Adhering to such guidelines in the field of psychometrics is considered by some as 'best practice' (Streiner *et al.*, 2015).

The COSMIN checklist has been adopted in this thesis as these guidelines were the most current at the time of designing this study. "*The aim of the COSMIN checklist is to evaluate methodological qualities of studies of the measurement properties of a PRO instrument, not for evaluating the quality of the PRO instrument itself" (*Mokkink et al., 2010, p. 545). The COSMIN checklist presents the relationships between all properties in a taxonomy of three domains: reliability, validity and responsiveness. The COSMIN checklist also allows for distinction between measures developed using statistical methods of measurement theory based on Classical Test Theory (CTT) or Item Response Theory (IRT) to be made. An explanation of each of these criteria is given throughout the rest of this section.

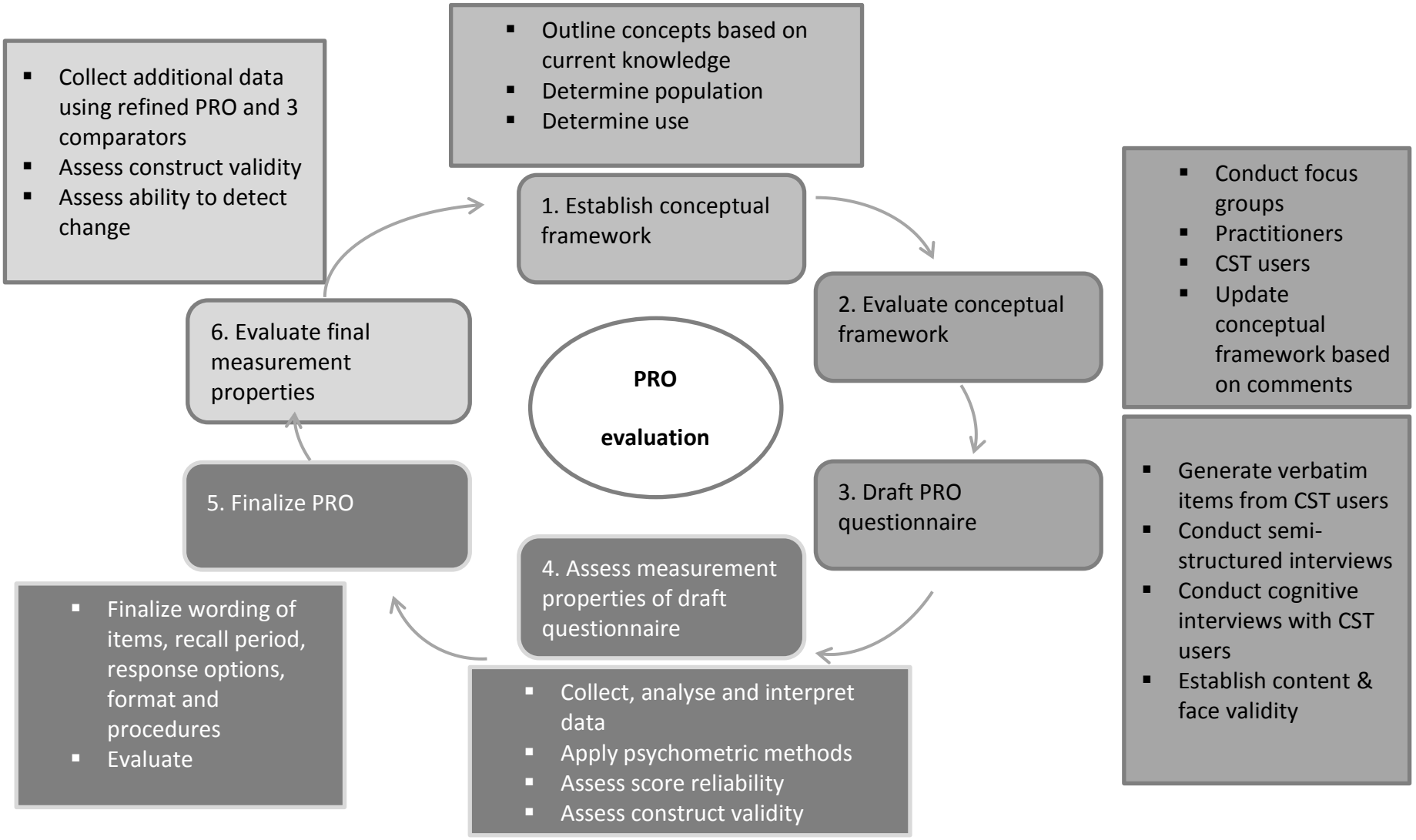
3.2 Mixed methods

Figure 6 has been adapted from the FDA Guidance (2009) and shows the sequential flow and iterative approach used in this study (see pg. 80).

In health psychology there is a tradition of developing new questionnaires using mixed methods approaches (Bishop, 2015). Mixed methods research

involves using quantitative and qualitative data (Creswell & Clark, 2007), allowing researchers to draw on the strengths of each.

Figure 6: Evolution of instrument development (adapted from FDA Guidance)



Mixed methods research can be viewed as a methodology or as a method, as Creswell and Clark (2007) explain:

"As a methodology, [mixed methods research] involves philosophical assumptions that guide the direction of the collection and analysis of data and the mixture of qualitative and quantitative approaches in many phases in the research process. As a method, it focuses on collecting, analysis and mixing both quantitative and qualitative data in a single study or series of studies" (pg. 5).

As in qualitative methods, mixed methods research is characterised by its iterative nature and the use of a cyclical approach (Tashakkori & Teddlie, 2010) is imperative.

According to Bishop *"sequential designs make it easier to retain the benefits of each method and to evaluate each method according to method-specific quality criteria, because only one study is completed before the other begins"* (2015). Other practical implications for a sequential design include:

- an accessible approach for a single researcher to develop the required research skills with training and support;
- it can take longer than other methods (gathering qualitative data, for instance interviewing one participant is more time consuming than inputting the data from a completed questionnaire);
- regulatory approvals require that each component is specified at the start of the study;
- findings can be published separately when each study is complete or together when the overall study is finished.

3.3 Theoretical perspectives

There are many ways in which to view the world and as both quantitative and qualitative methodologies and methods are to be used in this thesis, a brief outline of the theoretical perspectives and views which suit the context of this thesis are outlined. Quantitative approaches are associated with positivist (realist) beliefs in

an independent reality that is knowable, using objective measurement as an essential part of 'discovering' universal laws governed by behaviour (Bishop, 2015). A qualitative approach is underpinned by a relativist's position, believing that the world is knowable through conceptual frameworks. These frameworks may differ between individuals and cultures. Knowledge is understood *via* subjective means and is local and contextualised (Bishop & Holmes, 2013).

3.4 Qualitative methods rationale

This section discusses the methods an rationale behind adopting each of the methods used in this thesis.

3.4.1 Focus groups

Over the past two decades focus groups have been used increasingly in health care research to evaluate the satisfaction of clients with particular services (Parsons & Greenwood, 2000) and for promoting public health matters (Kitzinger, 1993). Focus groups have also been used to develop surveys or questionnaires (Nassar-McMillan & Borders, 2002) in order to find appropriate terminology for a certain population; applying either an unstructured agenda with the aim of generating hypotheses (e.g. item generation) or a structured agenda for those groups conducted to test hypotheses (e.g. item refinement). Open-ended questions are used as a means of gathering further information (Henderson, 1994). Krueger (1994) suggests that a minimum of two or three groups be held with groups sized between 4-6 participants thus, large enough to generate discussion, yet small enough to maintain adequate control over the agenda.

Focus groups - strengths

Focus groups are suggested in the FDA guidance (2009) as a method to ensure understanding and completeness of the concepts under discussion. In addition, focus groups make the most of communication between the participants in order to generate data and, therefore, use group interaction as part of the method (Kitzinger, 1995). Group interaction is useful for examining how people think and why they think the way they do, based on their knowledge and experiences.

While providing the researcher with rich experiential data (Morgan, 1988) focus group processes can help people to explore and clarify their views as participants exchange anecdotes and comment on each other's experiences and ideas about a pre-identified topic, here the topic of CST.

Kitzinger suggests that focus groups are used to reveal '*dimensions of understanding that often remain untapped by more conventional data collection techniques*' (1995). Nassar-McMillan and Borders (2002) imply that there are no definite rules for the use of focus groups, making their use a flexible method which can be adapted to suit the needs of the researcher and, therefore, focus group methods lend themselves to the aims of this research. According to Morgan (1988), careful planning and attention to detail are necessary if useful data are to be collected from focus groups.

Focus groups – limitations

The results gathered from a focus group study cannot be generalized due to the small number of participants and lack of random sampling (Stewart & Shamdasani, 1990). Yet, utilizing the experience of those who have experienced CST provides what Patton (2015) calls "information rich" cases. Focus groups can be time consuming due to the facilitation of data collection, transcription and data analysis and, therefore, expensive. Some participants can dominate the discussions if not facilitated well and focus groups are not such a useful method for individuals who are less comfortable speaking in a group setting.

3.4.2 Interviews

Interviews can be used to explore subjective meanings and issues that may be too complex to investigate through quantitative means (Banister *et al.*, 1994). There are many different styles of interviews. In this thesis both *semi-structured interviews* and *cognitive interviews* are used.

3.4.3 Semi-structured interviews

Semi-structured interviews rely on the interaction between the interviewee and interviewer and, although the interviewer has a topic guide, the interviews unfold in a conversational manner to allow the participant to explore issues that are important to them. The strengths of semi-structured interviews include: their informal and conversational nature, they allow for open responses in the participants own words, they can be carried out in a location convenient for the participant, and can be flexible for the interviewer to change questions to follow the direction of the conversation.

However, limitations can be like those experienced by focus groups. For instance, typically, only a small number of participants are involved, the sample may not be representative of the study population, the content may be difficult to compare as each interview is unique, and they can be time consuming in terms of data collection and analysis.

3.4.4 Cognitive interviews

Cognitive interviews were developed using theories of cognitive psychology as part of a multi-stage approach to questionnaire design (Drennan, 2002). They can be used to explore the intricacies of questionnaire design such as the use of irrelevant questions to respondents, the use of intrusive or insensitive questions and the number of potential problems in relation to understanding and successfully completing a questionnaire. Cognitive interviews allow the interviewer to find out how the responder has answered a question. Two techniques used for conducting cognitive interviews are 'think-aloud' and 'verbal probing' (Collins, 2003). Table 13 shows the key differences between the think-aloud and probing techniques as reported by Collins (2003, pg. 235).

Table 13: Key differences between think-aloud and probing techniques

Think-aloud	Probing
- Respondent-driven	- Interviewer driven
- Lower burden on interviewer as respondent does all the talking	- Lower burden on respondent, as respond to interviewer asks questions
- Can make the interview more difficult for the respondent	- Can make the interview easier for the respondent

The strengths of cognitive interviews can include: identifying problems with questionnaires such as misunderstandings and incomplete concept coverage, a useful method of identifying potential sources of measurement error such as problems of comprehension, processing or communication can be used to assess existing questions and to test proposed revisions to the original questions.

The limitations of cognitive interviews include: they cannot provide quantitative information about the impact an issue might have or provide quantitative evidence that revised questions might be better than the originals, they may discriminate against less communicative participants and are time-consuming (data collection, transcription and analysis).

3.5 Quantitative methods rationale

The quantitative methods in this study are mainly from psychometrics a branch of statistics. Psychometrics concerns the application of analytical methods to measure people's perceptions, beliefs and judgements about physical phenomena and this forms a cornerstone in the development of health measurement methods (McDowell, 2006).

"The purpose of a psychometric analysis is to establish the extent to which a quantitative conceptualisation has been operationalised successfully" (Hobart, 2009, pg. 2.). Older methods are underpinned by a theory called classical test theory (CTT), whilst more modern psychometric methods use item response theory (IRT) as their basis.

3.5.1 Classical Test Theory

Classical Test Theory (CTT) has its origins in the fields of education and psychology where the aim of measurement was often in the testing students (Hobart and Cano, 2009). Nunnally and Bernstein (1994) suggest that *"CTT views measurement as the determination of quantity or how much of an attribute is present in an object"* (pg.21). CTT is a strategy to measure constructs that are not directly observable, "*it is suitable for measurement of constructs that follow a reflective model"* (De Vet. *et al.,* 2011, pg. 19) and has five main assumptions (Hobart and Cano, 2009) as listed in Table 14.

Table 14: The assumptions underpinning classical test theory

1	Each person has an observed score, which is equal to their "true" score plus an error score.
2	If a scale is administered to a person an infinite number of times, the mean of their observed scores is equal to their true score.
3	Error scores and true scores are not correlated. Errors of measurement are not related to the observed score.
4	The error scores associated with two scales are uncorrelated.
5	The error scores on one scale are uncorrelated with the true score on another scale.

Limitations of CTT

Limitations include the inability to accurately determine the values of the true score (*T*) or the error score (*E*) (Hobart and Cano, 2009). Therefore, as these variables are unobservable for individuals, the assumptions underpinning the theory cannot be tested. Other limitations in this approach to psychometrics is that item difficulty and item discrimination are group dependant therefore, dependant on the sample from which they are obtained and will be influenced by the heterogeneity or homogeneity of the sample and their ability to complete the test. Scores are test dependent and CTT does not allow a prediction on how responders may score on an item. The standard error of measurement (SEM) around individual patients'

scores is assumed to be a constant value regardless of the person's location on the range of a scale (Petrillo *et al.*, 2015, pg. 32). This suggests that responders scoring at either end of the scale (floor and ceiling) are as precise as those scoring in the centre of the scale (following a normal distribution).

3.5.2 Item Response Theory (IRT)

Item response theory (IRT) also has its origins in the field of educational measurement and is a set of modern psychometric methods focusing on the relationship between a person's unobservable measurement of the underlying trait (latent variable) and the probability of responding to each of the response categories of a scale item (Hobart and Cano, 2009, pg. 10). IRT differs to CTT due to the inherent property of invariance of both item parameters and ability parameters (Hambleton & Jones, 1993) which are sample independent unlike CTT.

IRT models are used to measure a patient's ability (De Vet, 2011). The construct (ability) is usually represented with the Greek letter θ (theta). Guttmann scales are used and consist of multiple items measuring a single unidimensional construct. The items are chosen in such a way that they have a hierarchical order of difficulty. This is known as a 'deterministic' model; for example: in the case of assessing a patient's ability to walk, scale items are ranked from "easy" to "difficult" with the possible responses being 'yes' or 'no'. If the patient answers 'yes' to an item, this patient will score 'yes' to all the easier items and *vice versa*. Hence, the scale can determine the patient's ability, e.g. if a patient can run for 5 minutes, then they must also be able to stand. In practice, a true Guttman scale is very rare, hence IRT is based on the probabilities of the responses (de Vet *et al.,* 2011).

IRT aims to find the item response model that best explains the data. The models consist of item and person parameters.

3.5.3 Rasch models

Rasch measurement theory (RMT) uses a simple form of measurement model for a single latent trait. It assumes that the item locations (item difficulty) and a person's

score can be estimated independently of the test items from which they were calibrated and of the ability distributed of the sample (Hambleton & Jones, 1993). The Rasch model requires items with binary responses coded as: yes = 1 and no = 0.

If data do not fit the Rasch model, researchers will seek to understand why and, if necessary, remove data, re-collect data or re-conceptualise the construct for it to fit the model (Hobart & Cano, 2009).

3.5.4 Limitations of IRT/Rasch

IRT prioritizes the data and aims to find a model that best explains the data and RMT prioritizes the Rasch model and, if the data do not fit, the hypothesis will need to be revisited (Petrillo *et al.*, 2015). Thus, in both IRT and Rasch, models tend to be complex and model fit can be problematic. Both IRT and Rasch require an advanced level of mathematical understanding and unique software is required when adopting these approaches. Large sample sizes > 500 are required for IRT models.

3.6 Choice of model

To choose the correct technique to construct and analyse the new PRO questionnaire, it is essential to identify correctly if CTT or IRT are applicable to complementary and alternative medicine and CST. Due to the complex nature of CST sessions and the context in which they are undertaken, outcomes are non-hierarchical and do not follow a Guttman scale. Given the assumption from the conceptual framework that the items have no hierarchy, these properties are much better suited to CTT. Furthermore, Streiner and Norman (2008, pg. 327) suggest it is wrong to use IRT to construct indices of quality of life, symptom checklists, and other tools where the items themselves define the construct, rather than being a manifestation of an underlying trait or when the underlying construct is itself multifaceted and complex, as in the case of CST.

3.7 Reliability

Reliability is required for two different aspects of scale validation and both are described below. Firstly, internal reliability checks if multiple items of a questionnaire measure the same thing. This form of reliability uses item correlations to assess the homogeneity of the scales. Secondly, repeatability reliability relates to the extent in which a measurement is repeatable with the same individual, measuring the same attribute using different measures or by different people using the same measure of an attribute (Nunnally, 1978). In the COSMIN taxonomy, the measurement properties in the domain of reliability are: internal consistency, repeatability reliability and the related concept of measurement error. Definitions and how these aspects are assessed follow.

3.7.1 Internal reliability

Also known as internal consistency, internal reliability determines the extent to which the items making up the scale measure the same construct. This aspect of reliability evaluates the degree of interrelatedness amongst the items on a questionnaire. The variance observed by an instrument can be split into two components: 'true' variance created by the differences between subjects and 'irrelevant' variance created by other means such as measurement error. The proportion of the total variance in the measurements due to 'true' differences between subjects is expressed as the reliability coefficient. Hence, reliability coefficients will always lie between 0 and 1 and are only applicable to the specific population for which it was calculated.

3.7.2 Repeatability reliability

Repeatability reliability describes aspects of repeatability and the stability of measurements (Fayers & Machin, 2007). For example; if an individual with a specific condition completes a questionnaire on a Friday and is asked to complete the questionnaire on the following Friday and their condition has not changed, the measurement should produce similar values. Repeatability reliability is based upon analysis of correlations between repeated measurements, over time (test-retest),

by different persons on the same occasion (inter-rater), or by the same person on different occasions (intra-rater) (Mokkink et al., 2010).

3.7.3 Intraclass correlation coefficients (ICC)

Intraclass correlation coefficients (ICC) consist of a ratio of variances (De Vet *et al*, 2011). To have a measure of variance, scores are squared. In this study the formula proposed by De Vet, *et al*., (2011) have been used due to their association with the COSMIN taxonomy.

The variance components are:

σ_p^2 variance due to systemic differences between 'true' scores of patients

σ_o^2 variance due to systematic differences between observers

$\sigma_{residual}^2$ residual variance i.e. random error variance

Equation 1: ICC formula

$$\text{ICC} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_o^2 + \sigma_{residual}^2}$$

(De Vet *et al*, 2011. pg. 104)

Depending on the study and situation the variance components will differ (other ICC subtypes are available, ICC agreement and ICC consistency), some will be part of the measurement error and some will not. If absolute agreement between raters is important like, for example in a school exam: teachers agree the marks below or above a cut-off point for passing an exam prior to marking tests and absolute agreement is sought. If they are marking tests to identify the best 10 students, only consistency is relevant and the order in which teachers rank the students is of interest (De Vet, 2011).

3.7.4 Measurement error

It is normal for there to be a discrepancy between the numbers used to represent a concept being measured and the actual value of that concept. This is known as measurement error. Evaluating reliability is a way of revealing the amount of error inherent in any measurement. Ideally, measurement error is a small fraction of the range of any observations (Streiner *et al*., 2015). Measurement error alone is not informative. However, if measurement error is compared with the information about any expected 'variability' between clients being assessed, this would indicate the ability of the questionnaire to make a distinction between clients. Sources of error include variations within a test, variations between tests and systematic errors between observers (Nunnally, 1978). Multiple factors will influence how individuals respond to self-report questionnaires and the factors being measured, therefore, a score may be subject to a larger measurement error (Field, 2013).

3.8 Validating a PRO

Validity addresses the question 'does the instrument measure what it claims to measure?' Aspects of validity include content validity (face validit), criterion validity and construct validity which comprises of structural validity, hypotheses testing and cross-cultural validity.

3.8.1 Content and face validity

Content validity is an in-depth check to make sure that all known aspects of the domains of interest are covered by the measure. Face validity refers to the simple check that each item appears to measure the correct domain. Both aspects of validity can be assessed qualitatively in the context of cognitive interviews or focus groups that include experts and patients.

3.8.2 Criterion validity

Criterion validity identifies a 'gold standard' and explains how closely the scores of a measurement instrument reflects that standard (Mokkink, 2010).

3.8.3 Construct validity

Construct validity refers to the extent in which the scores of a questionnaire are consistent with hypotheses that the instrument correctly measures the underlying theoretical construct or domain (Mokkink, 2010).

3.8.4 Structural validity

Structural validity is the degree to which the scores of a PRO are an adequate reflection of the dimensionality of the construct to be measured (Mokkink, 2010). Structural validity is determined by identifying whether a construct exists of one or more dimensions (or factors), as this should be considered in further hypothesis testing (De Vet *et al*., 2011, pg. 169) (see pg. 117).

Factor Analysis (FA) is a versatile method which can be applied for different reasons. In its broadest sense, FA is used to summarise data so that relationships and patterns can be seen more easily (Yong and Pearce, 2013). FA attempts to capture the variance in the data through common factors (Child, 2006 cited in Yong & Pearce, 2013). Exploratory factor analysis (EFA) is used when there are no clear hypotheses about the underlying factors under investigation such as a means of reducing items on a novel questionnaire (Nunnally, 1967). A predetermined factor structure and *a priori* hypotheses can be tested to check whether data fit the theorized model using confirmatory factor analysis (CFA) (Floyd, 1995).

3.8.5 Convergent validity

Convergent validity evaluates correlations between two measures that are supposed to be measuring the same construct. The COSMIN checklist requires '*a priori* hypotheses' about possible relationships with the scores of other measures (Mokkink et al., 2010).

3.8.6 Cross cultural validity

Cross cultural validity addresses whether a measure is valid for use in a culture other than the sample in which it was originally validated. For example, during the

design phase, if a translation is required, have considerations been given to interpretations, to cultural relevance of items and to comprehension?

3.9 Responsiveness

Responsiveness addresses the question 'does the instrument detect change over time that matters to clinicians and patients?' The COSMIN checklist defines responsiveness as 'the ability of an instrument to detect change over time in the construct to be measured' (Mokkink *et al*., 2010b). In a research setting, evidence of a statistically significant change between observations is made at separate times when a change in the underlying construct is expected. An example of this would be the change in scores before and after an intervention. In this context 'clinical significance' would represent positive and noticeable improvements in the daily lives of those having a CST intervention. In contrast in a research setting, evidence of a 'statistically significant' change (ruled by the p-value and confidence intervals) would be seen between observations made at separate times when a difference of $p < 0.05$ is found. According to Fayers and Machin (2007) statistical significance does not imply clinical significance: *"statistical significance tests are concerned solely with evaluating the probability that the observed patterns in the data could have arisen purely by chance"*.

3.10 Interpretability

Interpretability refers to the meaning attributed to the scores of an instrument by patients, clinicians or researchers. Considerations are to be given to the following aspects:

- Distribution of the scores
- Floor and ceiling effects
- The availability of scores and change scores for relevant sub groups.

3.10.1 Distribution of scores of the instrument

The way data are distributed is important. A visual representation in the form of a histogram, which is a graph plotting values of the observations on the horizontal

axis with a bar showing how many times each value occurred in a data set, is a valuable way of evaluating the distribution of a data set. One reason why the distribution is important is because many statistical tests make assumptions about how data are distributed (Dancey & Reidy, 1999) with a 'normal' (Gaussian) distribution a frequent requirement. Understanding the distribution of the scale scores can also aid interpretation of the measurement properties of a scale (De Vet *et al.*, 2011) as it will aid in the identification of any edge effects.

3.10.2 Evaluating edge effects

A way of examining the range and distribution of responses is to check for floor and ceiling (edge) effects. Floor effects occur when a significant percentage of respondents score at the minimum score and the ceiling effects at the maximum score. It is an indicator that the scale may not cover the necessary ability range of the sample. If more than 15% of participants respond at either end of the scale, this can be rectified by generating extra items at the appropriate end of the scale during PRO development (McHorney, 1995). Floor and ceiling effects may be seen when an instrument is used in a new population. In a research setting, edge effects can also present problems in longitudinal analysis, since, if patients score at one end of the scale at baseline indicating that they are healthy, any further improvements will not be identified by the instrument (de Vet *et al*, 2011).

3.10.3 Interpreting change

Outcome measures can be used in research settings where 'population-based reference values' are utilised as well as in clinical settings where 'patient-orientated' methods (Fayers & Machin, 2007) and calculating change scores with pre and post-treatment evaluations are important in both settings.

In the development of a PRO for CST for use within clinical practice, two important benchmarks are required:

- The measurement error, expressed as the smallest detectable change (SDC)
- Minimal important change (MIC).

3.10.4 Smallest detectable change

Van Kampen *et al*. (2013) define the smallest detectable change (SDC) as "*a measure of the variation in a scale due to measurement error. Thus, a change score can only be considered to represent real change if it is larger than the SDC"* (2013, pg. 2). De Vet *et al*. (2011) emphasise the importance of using a measurement of the error such as the SEM (pg. 91) based on test-retest parameters and not on internal consistency measures such as Cronbach's alpha (α). This is because α is assessed at a single time point, not measured at different time points when a variation in the data is shown. Furthermore, the SDC is expressed using the same units as the original measure, hence their value in clinical practice.

3.10.5 Minimal important change

The minimal important change (MIC) is defined as the smallest measured change score that patients perceive to be important (van Kampen *et al.*, 2013). The two most common methods of determining the MIC are anchor-based and distribution-based approaches. Anchor-based approaches use an external benchmark to decide what patients, or their clinicians, consider as important improvements or important deteriorations (De Vet *et al.*, 2011). Distribution-based approaches focus on the distributional features of the sample, reporting the observed changes in the questionnaire under study to some form of variation to obtain a standardised metric (De Vet *et al.*, 2011).

3.10.6 Other considerations regarding interpretability

More general considerations to be given during the development or selection of an instrument include: how precise are the scores of the instrument? Is the instrument acceptable to patients? Is the instrument easy to administer and process? If an item is frequently missed, this could suggest an issue with the formulation or understanding of the item by the participants.

3.11 Chapter Summary

To summarise, this section introduced the methodological context of this thesis and the rationale behind adopting the mixed methods used in this study. The qualitative methods were described and quantitative measurement theories of CTT and IRT including the strengths and limitations of each model were presented. The psychometric properties by which PROs can be evaluated were introduced.

4 Methods

This section considers the ethics and governance for this study. Attempts to minimise bias are presented. Study design and a draft conceptual framework of CST outcomes are presented and the methods used to develop and evaluate the framework are described including the recruitment of participants, how data were collected, analysed and refinements made. A draft PRO is presented and the methods used to develop and assess the measurement properties and validity of the questionnaire are given.

4.1 Ethics and governance

4.1.1 Ethical approval

The University's Biomedical Research Ethics Committee (BREC) approved the ethics of this study on 20th May 2015, REGO-2015-1499 (See appendix 1). Ethics were considered in relation to the study design, interview process, the personal information provided and data gathered, and how data were stored during all methods undertaken in this thesis.

Travel expenses including mileage were offered to client and practitioner participants involved in the focus groups and lunch was also provided for each group. Participants were not paid for their time.

4.1.2 Rigour and validity

The COnsensus based Standards for the selection of health status Measurement INstruments (COSMIN) (Mokkink *et al.*, 2010) checklist has been applied to uphold methodological rigour in reporting this thesis. Consideration has also been given to any potential bias.

4.1.3 Bias

To minimize bias, the researcher's clients were not asked to participate in this study to avoid any conflicts of interest. Impartiality was maintained as not all practitioners who engaged in the recruitment process were known to the

researcher. Clients who took part in the focus groups and interviews were assigned an ID to give anonymity and clients also returned anonymised questionnaires during data collection. The research was supervised by academics that had no allegiance to the field of CST. This ensured that the research was scrutinised for bias throughout the duration of this study. Several ways in which bias was considered and minimized during questionnaire development included: a) end-aversion or central tendency bias which refers to the reluctance of some responders to rate at the extremes of a scale and b) ensure that participants have the appropriate answers to enhance optimising (the ideal way of responding to an item) and limit satisficing (giving an answer which is satisfactory, but not optimal) (Krosnick, 1991). The wording of the response options was planned to be evaluated with CST users to avoid such biases. To diminish 'yay-saying' bias, the tendency to give positive responses (Couch & Keniston, 1960), some items on the new PRO were planned to be reverse-scored to ensure that the responders did not give the same answer in a straight line down the page.

Strategies to enable reflexivity

To enable reflexivity, different strategies were adopted throughout the study including: academic supervision, personal psychotherapy, keeping a field work journal and scheduling time away from the study to process and reflect on the journey. These are discussed on pg. 212.

4.2 Study design for PRO Development

This study adopted exploratory, sequential mixed methods; building on the datasets during each step of the outcome measure design and validation process as described in FDA guidance (Figure 6, pg. 80). Briefly, this process was broken into a number of phases: developing a conceptual framework of CST outcomes based on previous qualitative work (Brough, 2012; Brough *et al.*, 2015) and the current literature, the evaluation of the conceptual framework with focus groups of both CST practitioners (experts) and CST users, constructing draft forms of the questionnaire and establishing content validity of the Warwick Holistic Health

Questionnaire (WHHQ), cognitive interviews with CST users to pre-test the questionnaire; and the use of quantitative methods to assess and evaluate the measurement properties of the WHHQ. All methods are described in detail in this section.

4.3 Establishing a conceptual framework for CST

This section explains the importance of the conceptual framework, describes the methods used to develop a hypothesised conceptual framework of CST outcomes, describes the population used in this study and states the purpose for which the PRO is to be used.

4.4 The importance of having a conceptual framework

To measure changes in outcomes after a course of CST, areas of importance as described by individuals before and after having a course of CST must first be identified. Without identifying all the key areas and concepts that make up CST, there is a risk of missing something important in a new questionnaire. This would make the new questionnaire unable, correctly to measure any improvements in health or wellbeing of those having CST.

4.5 Concepts of CST

The initial domains for the conceptual framework of CST outcomes are based on previous work (Brough, 2012) as described in section one (pg. 28) which involved interviewing n=29 CST clients and asking them about any changes they noticed and attributed to CST. Using data from these interviews, a thematic map of CST (see Figure 7, pg. 125) was created to present a visual overview of the hypothesised key areas of importance. Clients implied that health and wellbeing outcomes were at the core of their changes; with key conceptual domains of mind, body (physical functioning), spiritual wellbeing and social wellbeing.

To operationalise the key domains, these broad topics were then split into sub-domains to provide more detail about what each domain represented. The items which make up a domain are referred to as components in this thesis. The

components connected to the domains and sub-domains are the processes, actions and activities of how individuals relate to those aspects of their lives and will be reflected by items on the questionnaire, see Table 32.

Table 32 (pg. 169). As no other conceptual framework for CST is available, a conceptual framework for CAM has been reviewed (Verhoef *et al.*, 2006) to help make comparisons, see Table 49, (pg. 218).

4.6 Conceptual framework of CST outcomes and draft questionnaire assessment

4.6.1 Sampling strategy

Purposive sampling was adopted to ensure both CST practitioners and CST users were an integral part of the development and evaluation of the new WHHQ, utilising both experts and representatives from the population of CST users (Krueger & Casey, 1994). This sampling strategy was chosen based on the aims of the study (pg. 27). Group participants were not combined, to enable comparisons to be made between CST practitioners and CST clients.

4.6.2 Recruitment

A practical way of reaching CST practitioners was through CST practitioner groups. A multi-threaded recruitment strategy was adopted to recruit study practitioners through the CSTA. An email from the CSTA secretary was sent to members; information was put on the members-only pages of the CSTA website and an article was placed in the CSTA's quarterly membership magazine to notify members of the intended study. An introduction pack that included an outline of the project, a poster for practitioner clinics (see appendix 2) a patient information leaflet (PIL) (see appendix 3) and a consent form (for participation and recording) appropriate to each stage of the research process were developed and uploaded to the CSTA members' section of the website, so that practitioners could download them as required. Documents were updated throughout the research process. Expressions

of interest were made by practitioners and accepted *via* email or telephone. Practitioners who expressed an interest in any study stage were listed on a database.

4.6.3 Consultation

Consultation with practitioners beyond those who participated in the focus groups was key to the research process, since building relationships with them was important to the future use of the questionnaire. It was hoped that, by getting practitioners involved in different aspects of the planning and development of this study and getting consensus from practitioners about the comprehension of the questionnaire instruction documents, they would support the study when necessary.

4.6.4 Inclusion and Exclusion criteria

Practitioners:

Practitioners were recruited for the focus groups.

Inclusion criteria were:

- Good command and comprehension of the English language
- Able to travel to focus group location
- Able to give informed consent
- Aged 18 and over.

Exclusion criteria were:

- Practitioners from countries where English is not the first language were excluded from this study.

CST Users:

Client participation was required in the focus groups, cognitive interviews and questionnaire evaluations.

Focus group and cognitive interview inclusion criteria:

- Aged 16 or over (within practice, individuals over the age of 16 are considered as adults)
- CST users who had received at least four sessions of CST in the past
- To be receiving CST at the time of recruitment
- Ability to speak and understand English
- Ability to give informed consent.

Focus group and cognitive interview exclusion criteria:

- Clients who were receiving multiple treatment modalities during their sessions e.g. psychotherapy and CST, or acupuncture and CST
- Clients who may be traumatised and unable to complete a questionnaire
- Clients of researcher (NB).

Questionnaire evaluation: inclusion criteria

- new or existing client of recruiting practitioner
- aged 16 years and above in accordance with the ethical agreement
- first language of English and good comprehension skills.

Questionnaire evaluation: exclusion criteria

- previous inclusion in any part of the study.

4.7 Data collection

As this study followed a sequential mixed methodology, the process undertaken for each step is presented consecutively.

4.8 Practitioner focus groups

Written informed consent was given by all practitioners by completing the form sent *via* email or post, prior to the focus group. Participants were contacted by telephone or email to discuss the practicalities of taking part. This contact was followed up with a written letter outlining the focus group details including date, time and venue of group meeting and how to claim for travel expenses. The participants were asked to document their thoughts using the schema below (questions 1 to 9) based on open ended questions prior to the meeting so that they

could be used as prompts for discussion and debate during the session. It was requested that these documents be handed in at the end of the session for cross referencing with the transcripts and researcher's notes. A research assistant supported the smooth running of the focus groups.

A copy of the proposed working conceptual framework (see Figure 7, pg. 125) and an explanation about its development were provided to participants and the following questions were posed to the participants prior to meeting as a group:

1) What do you think about the four domains under evaluation? Domains include physical functioning, mental wellbeing, social wellbeing and spirituality.
2) How useful are the definitions in aiding your understanding of the domains?
3) What do you think about the components that make up the sub domains?
4) What do you like about the conceptual framework?
5) What don't you like about the conceptual framework?
6) In what way do you feel the working conceptual framework could be improved?
7) Based on your thoughts so far, how can the conceptual framework best be depicted?
8) Have we missed anything?
9) Make a note of anything else you may wish to raise at the focus group discussion in relation to the topic at hand.

4.8.1 Researcher's checklist

A checklist was made (Krueger & Casey, 1994) and consideration given to the following details prior to the focus group meeting: logistics, making sure the room was satisfactory, that background noise was minimal for recording purposes, name cards were prepared for participants, extra batteries, pens, copies of handouts, the list of questions for prompts, refreshments for lunch. Prior to each session, NB met with the research assistant to discuss topics of small-talk conversation that would

be appropriate as participants were arriving, ensuring that exchanges were kept 'off topic' until everyone had arrived and introductions had been made.

4.8.2 Facilitation of the focus groups

Ensuring ethical guidelines were adhered to (Rubin & Rubin, 1995), on meeting and prior to the start of the focus groups all consent forms were referred to and signatures checked and copies supplied. Participants were reminded that the discussion would be recorded and notes would be taken, that they could refuse to answer any question at any time and that they could withdraw from the process at any point without any explanation. The principal researcher (NB) informed the group that her role was to glean information from the group rather than to provide "correct information" (Nassar-McMillan & Borders, 2002) and to ensure the group remained on task and within the time allowed. The research assistant was introduced and the participants were told that her role was to ensure the discussion was recorded and to document important points for summarising towards the end of the session.

4.8.3 Schedule for focus groups:

Schedule 1: Practitioner focus groups

Section	Moderators/Participants Moderator: NB Research Assistant: SM	Time allocation
Introductions Thanks for coming Health & Safety, house rules. Lunch - to be eaten as we go Recording - ground rules, speak clearly, like to hear from everyone; aim of session. Any practical questions.	Moderators Tent cards on seats with names	5 minutes
Opening questions Tell us your name, where you practise and what you most enjoy when you are not practising CST. (starting with ...and working clockwise). Are there any words of phrases you do not understand in the documents you were sent?	All - SM to make a diagram of who is sitting where?	10 minutes
General questions 1. What do you think about the four domains under evaluation? 2. How useful are the definitions in aiding your understanding of the domains? 3. What do you think about the components that make up the sub domains?	SM to make notes of main points	30 minutes
Key Questions 4. what do you like about the conceptual framework? 5. what don't you like about the conceptual framework? 6. In what ways could it be improved?	SM to make notes of main points	50 minutes
Response Shift 1) Recalibration of respondent's internal standards of measurement. 2) reprioritisation of respondents' values. 3) reconceptualization of the target construct.	Explore does this process sound familiar	15 minutes
Summary Have we missed anything? Any questions?	SM to write summary on flip charts.	10 minutes
Closure, Thanks, End.		

4.8.4 Debriefing notes

Debriefing notes were made after each session and observations and thoughts from both the principal researcher and research assistant were documented using the questions below to capture the important aspects of the discussion.

1) What were the themes?

2) What are the most important points learnt from this group?

3) What was surprising or unexpected?

4) What quotes were particularly helpful?

5) Does anything need to change before the next group?

Focus group sessions were audio-recorded and transcribed verbatim and participant names were codified to ensure anonymity.

4.8.5 CST users' focus groups - evaluating a working conceptual framework of CST outcomes

The process and documentation described on pg. 104 was also followed with CST users. Adjustments were made to the location and venue of this group to access 'local' users of CST. Refinements were made to the conceptual framework handout based on the findings of the focus groups undertaken with CST practitioners. These changes are reported in the results section of this thesis, see Table 16, (pg.133).

4.9 Draft PRO questionnaire

In this section, the methods used to draft the PRO questionnaire are presented, including the planning of instrument usage, naming the questionnaire, generating verbatim items from CST users, constructing a draft version of the questionnaire, conducting semi-structured interviews to determine item selection, conducting a consensus meeting with CST practitioners to check item selection, undertake cognitive interviews with CST users to pre-test draft questionnaire and assess face and content validity.

4.9.1 Determined population

The population used to support the development and evaluation of a new PRO are CST users over the age of 16 (in practice, individuals age 16 years and over are considered adults).

4.9.2 Determined use

The new PRO is to be evaluative, assessing changes in health and wellbeing of those having CST. The PRO is intended for use in a clinical setting.

4.9.3 Naming the questionnaire

Factors accounted for when considering the name of the new questionnaire were as follows: the questionnaire may be applied in other CAM settings, therefore, caution was given not to limit the potential use of the questionnaire by putting 'CST' in the title. Adopting the name 'Warwick', utilising the reputation and the brand of the University, was discussed. In previous work (Brough, 2012) users attributed CST as having effects on aspects such as mind, body, and spirit, as well as being holistic. The researcher wanted to incorporate this into the title, resulting in the title Warwick Holistic Health Questionnaire (WHHQ). It is commonplace to use acronyms to shorten the name of a questionnaire. The newly created and named questionnaire will be referred to as the WHHQ throughout this report.

4.9.4 Construction of a draft version of the questionnaire

A draft version of the WHHQ including details such as the layout, instructions on completion, time frames, response options, and the wording of each item was prepared at the same time as the conceptual framework. The starting point for this thesis was to review the previous work by Brough (2012) based on interviews with a large number of CST clients. These data were used to create a list of potential items for a PRO. This was done by using verbatim statements to create items when possible. For example, when participants reported "feeling more relaxed and calm" (Brough *et al*, 2012, pg. 170) an item was generated "I've been feeling relaxed and calm" from the interview data. This process was repeated creating items related to

physical body (functioning, symptoms, daily activity), mental health (mind, emotional affective, cognitive), spiritual wellbeing (awareness, relating to self and others), and holistic wellbeing. At that time the debate about the role of spirituality in healthcare and how to evaluate its influence on health was emerging and questionnaires which covered outcomes of this domain were considered to aid the item generation for this domain including the World Health Organization's Quality of Life - Spirituality, religion, personal beliefs instrument (WHQOL SRPB) (WHOQOL Group, 2006), JAREL spiritual wellbeing scale (Hungelmann *et al.*, 1996), The Self-Perception and Relationships Tool (S-PRT) (Atkinson *et al.*, 2004), (Atkinson *et al.*, 2004; Group, 2006; Hungelmann *et al.*, 1996; Ng *et al.*, 2005; Peterman *et al.*, 2002). An initial list of 35 items was created for the WHHQ.

When the conceptual framework had been evaluated and CST users and practitioners had given their input and feedback on the content, further items were generated. The conceptual framework was used as a map and items were generated for inclusion in the domain of social wellbeing and engaging in life; as were items covering the development of self-awareness, taking responsibility for self and everyday life, including life satisfaction and doing things I enjoy. This resulted in a list of 73 items for inclusion in the draft PRO. Attention was given to the way in which the items were phrased to keep the integrity of the participants' input and to be able to evaluate how the development of awareness changed as individuals' experience of CST deepened. Looking at the content of questionnaires in the field of mindfulness (Cardaciotto *et al.*, 2008) and life satisfaction (Bussing *et al.*, 2009; Bussing *et al.*, 2007) assisted thinking about items linked to this domain, though the reliability and validity of these items were not checked. Items created to evaluate symptoms were reverse scored to ensure the flow of completion was smooth, but also as a means of checking that responders were paying attention to how they answered the items.

As described in the literature review (pg. 60) many questionnaires were studied and the design, layout, instructions and response options observed. The initial WHHQ draft was, therefore, based on these observations. These steps enabled content and face validity to be assessed later using cognitive interviews.

4.9.5 Evaluating the draft PRO

As the geographical spread of participants was broad, practical and financial reasons led to the decision to undertake semi-structured interviews to facilitate the next phase of the research instead of the focus groups as initially intended. Therefore, each face-to-face semi-structured interview was carried out at a place convenient to the participant, providing it was suitable for recording purposes. The semi-structured approach was adopted to explore areas where the participants perceived gaps and difficulties in the proposed questionnaire and to tailor the questions to the position and comments of the interviewee.

Recruitment was undertaken as previously explained (pg. 100) via the CSTA. Participants were contacted *via* email or by post at the request of the participant, and an introductory letter which outlined the aims and objective of this research was sent with a draft questionnaire containing 73 items (see appendix 5). An outline led the participants through a series of prompts (see below), providing space for them to record their thoughts and opinions, responses were used as reminders or prompts for discussion during the interview and handed in to the researcher for cross-referencing purposes during analysis.

Prompts used to assist in constructing the questionnaire consisted of:

- Layout
    1. What do you think about the design and layout of the questionnaire?
    2. What do you like?
    3. What don't you like?
    4. Does anything need changing?
- Instructions
    5. Are the instructions for completion clear?
    6. Could you fill in this form without help?
- Recall period
    7. The questions focus on your symptoms over the last two weeks, is this a good time period?
    8. Should it be longer or shorter?

- Response options
    9. What do you think about the different response options?
    10. Can you understand them?
    11. Are there enough options?
    12. If you don't like them what would you like to see instead?
- Questionnaire name
    13. What do you think about the name: "Warwick Holistic Health Questionnaire"?

Participants were asked to select items from the list and to record the item number in one of six boxes, each box had a different heading: 1) items relevant to my experience, 2) I don't understand the statement, 3) the response options are not suitable, 4) items are repeated, 5) item addresses more than one point, and 6) items you would like to remove, explain why? Participants were asked to make a note of anything else they wished to raise at the interview in relation to the topic. The data in relation to each item were documented under the relevant topic and items selected by each participant were recorded on a spreadsheet for later analysis.

4.9.6 Data analysis for semi-structured interviews

All interviews were audio recorded, the data were transcribed, anonymised and the transcripts were checked against the audios. Inductive thematic analysis (Braun & Clarke, 2006) was used to identify, analyse and report the themes found in the interview data. Concerning the 73-item WHHQ, the number of times an item had been either selected or not selected for inclusion on the questionnaire and the participant's rationale for this, were considered as part of the analysis.

4.9.7 Consensus meeting with CST practitioners

An opportunity arose to consult with an established group of CST practitioners (n = 16), and it was decided that this would provide an opportunity to obtain practitioner feedback on the draft questionnaire. All practitioners consented to taking part, to adhere to confidentiality and to having the session recorded for

transcription purposes. The group was separated into four, given a copy of the draft 73-item WHHQ and allocated a domain listing the proposed sub-domains and items for that domain. Practitioners were asked to work through the outline described in the previous section (see pg. 109) to evaluate the design, layout, instructions, time frames, response options and the clarity of each items in the domain allocated to them. Participants documented their discussions and feedback on the forms provided. Each group was allocated 20 minutes to undertake the tasks and each group fed back to the wider group and questions or concerns were worked through. The group discussion was audio-recorded and each group returned the completed feedback forms. Data were collated from the feedback forms allocated to each group, sorted and analysed based on the feedback of each small group and the consensus of the wider group. The audio recording was listened to but due to the competing group discussions it was not possible to capture data in this way. Comparisons were made between this group and the individuals who took part in the interviews. The results of both the semi-structured interviews and consensus meeting with practitioners was used to determine the content and items on the questionnaire. As a result, the 73-item WHHQ became a 52-item WHHQ.

4.9.8 Pre-testing the draft PRO (52-item WHHQ)

Cognitive interviews were undertaken to assess the content and face validity, that is, to find out if the questionnaire was user friendly, easy to understand, and to evaluate if the language, format and response options were acceptable to participants. It also provided the opportunity to ascertain the average time it took participants to complete the 52-item WHHQ. Participants were interviewed at a convenient location for them. Consent was confirmed and an overview of the intention of the interview was given. The researcher followed an interview schedule (see appendix 6). Interviews were audio-recorded and the length of time it took for each participant to complete the WHHQ was timed with a digital stop-watch. The researcher remained quiet until an issue arose and the respondent was then asked to think aloud questions such as '*How did you go about answering that question?'* or '*How easy or difficult did you find this question to answer?*' Probing

was used on occasions when the interviewer noticed that the respondent hesitated before answering a question, revisited a question or did not complete a question. The audio-recordings were transcribed, and checked against the data.

4.9.9 Data storage

To ensure confidentiality, the Data Protection Act (1998) was adhered to and anonymisation or pseudonymisation of data was undertaken. Electronic data were backed up to an external hard drive and stored securely. Only the research team of the academic supervisors and researcher had access to the data. All data held on paper were kept under lock and key with only the researcher having access to it directly. Names and interviewee data were stored separately. As transcription of the focus group discussions was undertaken by a contractor, contracted transcribers were asked to sign an agreement to observe confidentiality of the data.

4.10 Testing the draft questionnaire in a group of CST users

This section describes the aims of this part of the study; the recruitment of participants, the inclusion and exclusion criteria, data collection and data handling, the chosen measurement model, the psychometric methods applied to the data, and analysis for assessing the measurement properties according to the FDA guidance.

4.10.1 Assessing the measurement properties

The aim of the sub-study was to field test the draft 52-item WHHQ with a group of CST users, evaluate its psychometric properties and reduce and refine the items.

4.10.2 Recruitment

CST users were recruited for this sub-study using the same methods reported above (pg. 100). CST practitioners registered to be involved *via* email and by return were allocated a practitioner ID (PID) and were sent a guide on how to administer the data collection. The guidance provided the aims and objectives of the study, the inclusion and exclusion criteria, the timelines for the study, how to administer the

data collection, including how to allocate a client ID (CID), and what information to document. Guidance on how to add scores on a completed questionnaire was given and how to obtain more questionnaire booklets, if needed. Questionnaire booklets and prepaid envelopes were sent by post to registered practitioners. Recruitment posters were displayed within practitioner clinics and CST users were invited to complete a questionnaire before a session of CST.

4.10.3 Inclusion and exclusion criteria of CSTA participants

Any new or existing clients of participating CSTA member practitioners were eligible for this study. CST users under the age of 16, those having multiple treatment modalities during their sessions e.g. psychotherapy and CST or acupuncture and CST, those who may be traumatised and unable to undertake the task at hand and anyone about whom practitioners had doubts, for whatever reason, were excluded.

4.10.4 Data collection and method of administration

The draft instrument (52-item WHHQ) was self-administered by CST users whilst waiting for their CST session or as part of their therapy session depending on the facilities and/or preference of each study practitioner. Participants were deemed to have consented by returning a completed questionnaire.

4.10.5 Data handling and cleaning

Completed questionnaires were returned by post to the researcher. On receipt questionnaires were checked to ensure they had correct PIDs and CIDs to enable tracking. Data were put into the statistical software package SPSS (IBM, version 22). To ensure confidentiality, the data protection act (1998) was adhered to. Data were stored as described in the data storage section (pg.112).

4.10.6 Scoring rules and missing data for WHHQ and comparator measures

To meet the design requirements of the COSMIN checklist considerations for scoring rules and for missing data were determined prior to data collection. Here all measures (including 52-item WHHQ and 25-item WHHQ, HEHIQ, SF-12v2 and

WEMWBS) and the way they had been scored and missing data handled are described.

WHHQ: Items were scored as indicated on the draft instrument, with each option given a rating from 0 to 4.

25-item WHHQ: A total score is calculated by adding the 25 individual statement scores. The minimum score is 0 and the maximum score is 100 with high scores implying higher levels of wellbeing.

Each questionnaire was scored according to the questionnaire instructions:

SF-12v2: Calculated as reported in the handbook (Ware et al., 2010). The physical component summary score yields a single score that can be used as an overall measure of physical health and the mental component summary an overall assessment of mental health. Scores of less than 40 indicate impaired functioning or wellbeing associated in the domain. The scores of each domain are not combined, the SF12v2 does not have a total score (Ware *et al.*, 1995).

WEMWBS: Items were added to create a total score between 14 and 70. Higher scores indicate higher levels of mental wellbeing.

HEHIQ: Each of the five subscales has a score and these are added to create a total score. Scores can range between 20 and 100, higher scores indicate more positive state of health.

Missing items were handled in the following manner: for each instrument, scores were mean imputed if the number of missing items on the instruments for a respondent was three or fewer, excluding demographic data and anchor questions; i.e. the mean score of the reported items was assigned to each of the missing items for that person. Scores were rounded to the nearest whole number.

According to Streiner and Norman (2008), if the number of omitted items is small (less than 5%), then assigning the mean score probably will not distort the results too much. In the case of WHHQ, 5% of 52 items is 2.6 items rounding up to 3.

If the number of missing items for a respondent was greater than three for any particular questionnaire, the total score was deemed to be missing and the respondent's data were removed from the dataset of that comparator measure.

Imputed datasets were analysed and reported, noting the number of frequencies of missing data per item. If a respondent gave two answers for an item or endorsed a point between two options, the mid-point between the two scores was inputted. For example: for a respondent endorsing both 3 and 4 for an item; 3.5 was inputted.

4.10.7 Floor and ceiling effects

Measures of central tendencies were determined and the standard deviation was calculated for each item. Histograms were created to provide a visual representation of the distribution. Items were analysed to identify any floor and ceiling effects at scale level and at item level to see if all response options were being used. No floor effects or ceiling effects were identified.

4.11 Measurement model

Classical Test Theory (CTT) was the chosen measurement theory that has been applied to this study. Details of the rationale can be found on pg.86.

4.11.1 Descriptive statistics

Exploratory data analysis was carried out to establish the demographics and characteristics of the sample. Measures of central tendencies were calculated, in this case, the mean due to the assumption of normality. The standard deviation was determined and graphical techniques to illustrate the data with frequency histograms being prepared for each item to assess the way the data are distributed. Floor and ceiling effects were also evaluated.

4.11.2 Psychometrics

Reliability

Reliability is the degree to which a PRO produces stable and consistent results.

Internal Consistency

Cronbach's alpha was calculated as a function of the number of test items and the average inter-correlation among the items. An $\alpha < 0.70$ would suggest that the scale is poor, an $\alpha > 0.9$ would suggest there are too many items in the scale and some may be considered for deletion (De Vet *et al.*, 2011).

4.11.3 Intra class correlation coefficients

A two-way mixed model of absolute agreement was investigated assuming that the errors are coming from both the patients in a random sample and the questionnaire. The intra-class correlation coefficient (ICC) formula for $ICC_{agreement}$ is as follows:

Equation 2: ICC absolute agreement

$$ICC_{agreement} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_o^2 + \sigma_{residual}^2}$$

Here, $\sigma_o^2 + \sigma_{residual}^2 = \sigma_{error}^2$ = total error variance. See pg. 61 for further details.

4.12 Measurement error

4.12.1.1 Standard error of measurement (SEM)

The standard error of measurement is a parameter of measurement error and is calculated by the following formula:

Equation 3: Standard error of measurement

$$SEM = SD\sqrt{1 - r}$$

Where SD = Standard deviation of the scores, r = reliability coefficient, such as Cronbach's alpha (α).

4.13 Structural validity

4.13.1 Exploratory factor analysis

In this thesis, exploratory factor analysis (EFA) was used to explore the factor structure. EFA was also used as an item reduction technique, see pg. 117. In this study, EFA was carried out using SPSS (IBM, version 22).

Extraction methods

There are many factor extraction methods (Yong and Pearce, 2013) three common methods include: maximum likelihood (ML), principal axis factor (PAF) and principal components (PC). ML is used to estimate the factor loadings for a population and is therefore, more useful for confirmatory factor analysis. PAF is used when the data violate the assumption of multivariate normality (Osborne and Castello, 2009). PC is the best method for reducing variables as it explains the most variance for any set number of factors (Nunnally, 1967) and was used in this study.

Rotation methods

Rotation methods are used to simplify and clarify the data structure, with each item loading on as few factors as possible. Two rotations types are possible, orthogonal and oblique. An orthogonal rotation is undertaken when it is assumed that factors are uncorrelated (Gorsuch, 1983) and an oblique rotation when the factors are

assumed to correlate. In this study, an oblique rotation was applied using the promax technique, where the loadings were raised to the power of four, or Kappa = 4. This is considered to be a good trade-off between the correlations amongst the factors and a simple factor structure (Gorsuch, 1983). Changing the Kappa value alters the amount the rotation procedure "allows" the factors to correlate (Osborne and Costello, 2009).

Factors

Factors are any linear combination of the variables in the data matrix and, after a factor is obtained, scores can be correlated with scores on each of the individual variables in the data matrix (Nunnally, 1967). Factor loadings refer to the factor variable correlations (Nunnally, 1967), meaning that the relationship of each variable to the underlying factor is expressed by the so-called factor loading (Grace-Martin, 2017). The size of the factor loading was checked when analysing the data to see if the item was explained sufficiently by the factor. To do this, an absolute loading cut-off point of ≥ .32 was set, and only items greater than this were considered to be associated, or "loaded" onto that factor. To check the quality of the EFA model, the item communalities were considered. Item communalities show how much of the variance in the variables has been accounted for by the extracted factors (Chetty, 2015). Item communalities >.80 were considered high and items such as these were deleted. Items with communalities <.40 indicate that the item is not related to the other items or that an additional factor needs to be explored. Cross loadings items occur when an item loads (≥.32 or higher) on two or more factors (Osborne and Costello, 2009) and this may suggest that items are strongly correlated. A decision is then made whether such items be dropped from the analysis. This depends on the data, and if the nature of the variables under analysis is complex.

The number of factors to be retained was determined *via* the use of eigenvalues of each factor and a scree plot. A scree plot shows eigenvalues and factors; factors to be retained are those which sit above the point of inflexion and which have an eigenvalue greater than one (Kaiser, 1960).

4.13.2 Model refinement

Once an initial model was generated, the results were analysed using scree plots and by evaluating eigenvalues as described on page 93; item communalities were assessed as described above, to identify item performance, such as items cross loading on one or more factor. The "worst" performing item was then deleted using the following criteria:

- Items with communalities > 0.8 were removed
- Items with absolute loadings < 0.3 were removed. If multiple items had the same loading value, only one was removed
- Items that were not loaded on any of the factors (free standing) were removed.

Once refinements were made, models were re-run using the same criteria until the items created a suitable factor structure and no items fit the removal criteria.

4.13.3 Assessing content and face validity

After EFA and refinement of the 52-item WHHQ to 25 items, further consideration was given to content and face validity. The qualitative results were revisited to see if items important to CST users and CST practitioners had been maintained. Items which had been deleted during the EFA refinement process and that reduced face and content validity were listed, see Table 29, pg. 165. At a meeting with CSTA member practitioners (n=60) in March 2015 a poll was undertaken to gain opinions about these items, practitioners were asked to consider each item based on the criteria below:

Each item deleted by EFA refinement was re-evaluated against the following criteria:

- Does the item add anything to the questionnaire?
- Will including the item influence response rates?
- Will excluding the item influence response rates?
- Might the item be sensitive to change over time?

Items were reinstated based on a majority vote of practitioners who took part in the poll.

4.14 Evaluating the measurement properties of the final version of the WHHQ

4.14.1 Construct validity

Three of the eight questionnaires reviewed in this thesis were selected to use as comparator measures for evaluating convergent validity. These were the HEHIQ as it was developed with a population similar to that of CST users and its content matched most of the areas identified as important to CST users, the WEMWBS which unlike any of the other questionnaires, measures mental wellbeing, an important outcome of CAM use (Verhoef *et al.*, 2006) and is well validated. Also well validated, the SF-12v2 was adopted as it captures some of the domains of interest and is well respected with orthodox medicine and health care policy makers.

4.14.2 Hypothesis testing

It was hypothesised that the constructs under evaluation positively correlated with the three comparators (WEMWBS, SF-12v2 and HEHIQ) against that of the 25-item WHHQ. For exploratory purposes, the following hypotheses were proposed:

- mental wellbeing domain positively correlates with the WEMWBS and the SF-12v2 MCS score.
- physical functioning domain positively correlate with the SF-12v2 PCS score.
- spiritual and social functioning domains positively correlates with the HEHIQ.

4.14.3 Strength or magnitude of the relationship

Pearson's correlation coefficient (*r*) was used to measure the strength of the relationship between two measure scores (Dancey and Reidy, 1999).

This varies from -1 to +1. The strength of the relationship is reported using the following cut points as a guide: 0 none, weak ≤ 0.30, moderate ≤ 0.60 and strong ≥ 0.70.

Overall, it is hypothesised that a strong correlation of ≥ 0.70 will be seen in a positive direction on all the above.

4.14.4 Bland and Altman plots

A visual check of agreement between two questionnaires was carried out using Bland and Altman plots (1986). Bland and Altman (1986) advocate the use of graphical methods to plot the different scores of two measurements against the mean for each subject.

4.15 Assessing the 25-item WHHQ's ability to detect change

4.15.1 Responsiveness

Assessment of the ability of the 25-item WHHQ to detect change was done by taking two measurements, one before the first session of CST and the second before a follow-up session of CST between one and four weeks apart. A global rating scale was used during the follow-up session:
"In between completing these questionnaires, the first and second time (this occasion) my health has:"

a) changed a lot for the worse
b) changed a little for the worse
c) stayed the same
d) changed a little for the better
e) changed a lot for the better

4.16 The smallest detectable change (measurement error)

The SDC was calculated using the following formula. The formula for SDC the change is larger than the measurement error (SEM).

Equation 4: SDC

$$SDC = d \pm 1.96 \times SD_{difference}$$

Here, d= effect size or Cohen's d; $SD_{difference}$ = standard deviation of the differences

Or in the absence of systematic difference, larger than the SEM

Equation 5: SDC 2

$$SDC = \pm 1.96 \times SD_{difference} = \pm 1.96\, x \sqrt{2} \times SEM$$

4.17 Minimal important change (MIC)

Minimal important change (MIC) is about clinical benchmarking. An anchor-based approach was used to identify how CST users rated their overall wellbeing prior to completing the 25-item WHHQ at both baseline and follow-up. "How do you rate your overall wellbeing today?" was presented and the response options included; poor, fair, good, very good and excellent. A second anchor question was asked prior to the follow-up session of CST. The MIC is defined as the mean change score in the subcategory of patients whose health status 'changed a little for the better' in accordance with the anchor question described on page 121.

The distribution-based method usually considered in health measurement, is Cohen's d, or effect size (Streiner *et al.*, 2015).

Equation 6: Cohen's d

$$\frac{mean\ change}{SD\ at\ baseline}$$

4.18 Chapter summary

This section described the introduction of the study design in addition to the methods used for the development and evaluation of the conceptual framework of CST outcomes. A draft item list of 73 items for inclusion in the 25-item WHHQ was provided and the process by which it was derived was described. The measurement model and psychometric methods used in the evaluation of the WHHQ were also presented. The way in which the questionnaire was refined based on the data analysed is described as well as the way reliability and construct validity were

assessed. The modifications undertaken are reported in the results section (pg. 145).

5 Results: Evaluating the conceptual framework of CST outcomes

5.1 Introduction

This section presents a Conceptual Framework of CST outcomes and summarises the focus group discussions undertaken to evaluate the conceptual framework. Refinements made to the conceptual framework based on the focus group discussions are reported. Figure 7 (overleaf) shows the initial version of the framework which was developed over the course of 6 months using the literature (Brough, 2012) and expert opinion. The content of the working conceptual framework (see Figure 7) included the domains of spirituality (including being present, connecting to self, divine and nature, trust, faith in life and having compassion); physical functioning (involving mobility, posture and functioning, daily activities, sleep and symptoms); mental wellbeing (consists of mental outlook, emotions/feelings, self-concept, self-agency, self-efficacy and self-care); social wellbeing (including engaging in life which was made up of 'doing things I enjoy' and 'feeling connected to my friends and family'). Definitions related to each domain were prepared to aid in interpretation (see Table 15).

Appendix 4 lists these revisions and an A2 copy final conceptual framework of CST outcomes can be found at the end of this thesis.

Figure 7: Working draft of Conceptual Framework of CST outcomes

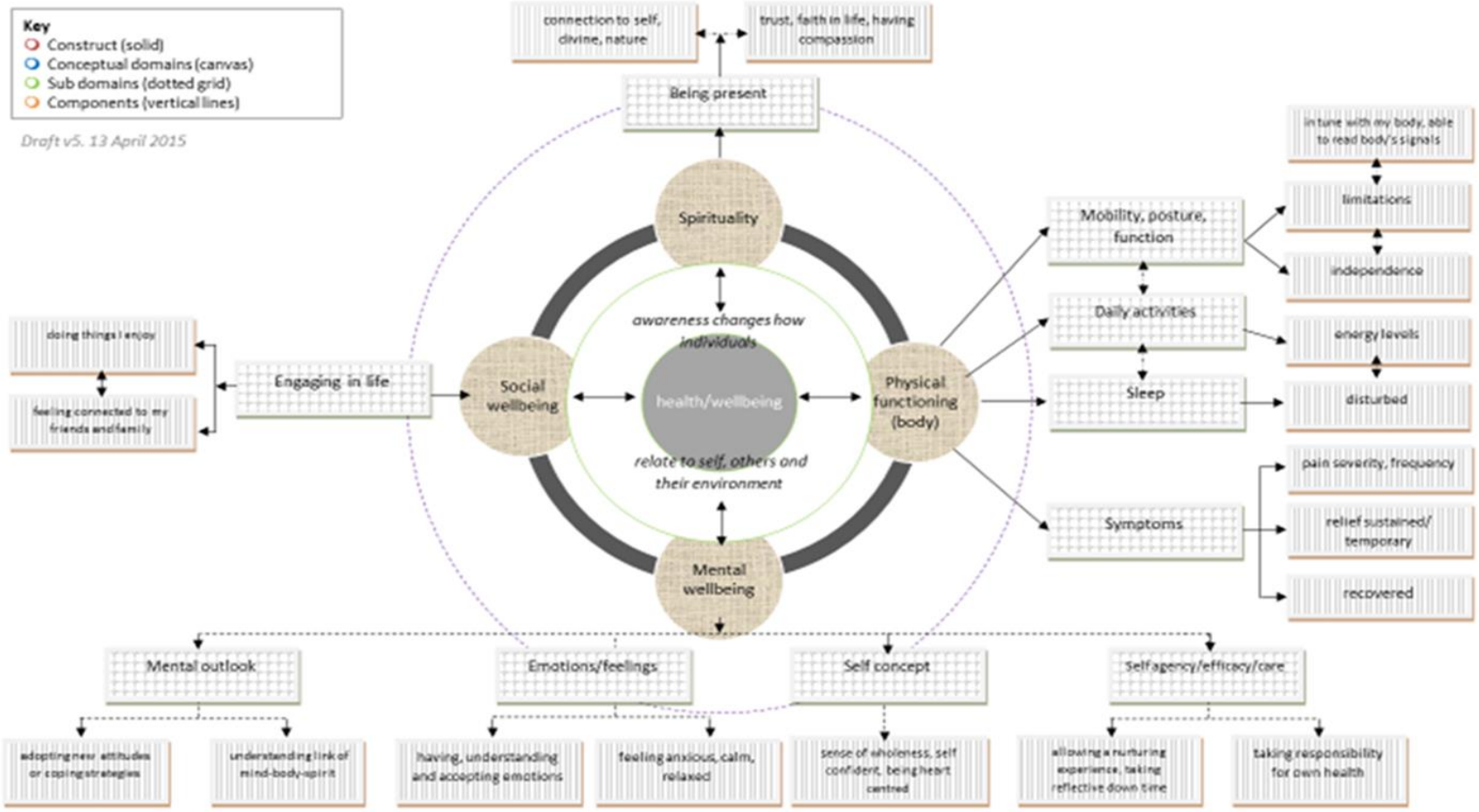


Table 15: Conceptual Framework of CST Outcomes Domain Definitions

Health	A state of complete physical, mental and social wellbeing not merely the absence of disease (WHO, 1946). Here, health covers a spectrum from disease to wellbeing.
Wellbeing	Covering physical, mental, spiritual and social domains including a sense of vitality, undertaking activities which are meaningful and engaging, feelings of competency and autonomy, personal development good relationships with others, inner resources to aid coping when things go wrong and resilience to changes that are beyond one's immediate control (adapted from NEF, 2009).
Spirituality	Includes personal and internally focused elements such as core values, philosophy and meaning of life or the way one conducts life. It is the search for the sacred or divine through any life experience (Mytko and Knight, 1999) and connectedness to self, nature and the wider universe.
Body (physical functioning)	This is a spectrum represented by physical symptoms that may influence posture, inhibit function and mobility. With optimum physical functioning at one end of the continuum with ability and daily activities and how they are affected (poor physical function) at the other end of the continuum (Brough N., *et al*. 2014).
Mental wellbeing	Feeling good and functioning well, covering both psychological and emotional wellbeing, attitudes and outlook on life (FPH, 2010).
Social wellbeing	A sense of relatedness and connectedness to other people (NEF, 2009).
Being present	The state of being attentive to and aware of what is taking place in the present (Brown and Ryan, 2004).
Awareness	The state or condition of being aware; having knowledge: stems from the word 'aware' refer to an individual sense of recognition of something within or without oneself.

5.2 Focus groups with CST practitioners

Approximately 500 CST practitioners were invited to participate in the focus groups to evaluate the working draft conceptual framework. The intention was to recruit

practitioners with varying levels of experience, working in diverse settings throughout the UK. Seven practitioners consented and took part. Further details can be found in the methods section on pg. 102.

5.2.1 Demographics of focus groups 1 and 2

Two practitioner focus groups were held. The first comprised of four practitioners and took place in Oxford in June 2015. It lasted for 1 hour and 51 minutes. The second comprising of three people, was held in London in July 2015. It lasted for 1 hour and 23 minutes. The participants in group one were all female, had been practising CST between 1 and 4 years and were from Leicestershire, Surrey, West Sussex and Swindon. All but one worked in private practice, with the remaining participant, a physiotherapist, working within the NHS using CST when appropriate. The participants in group two (two females and one male) had been working in private practice between 10 and 13 years and were based in London, Gloucestershire and the Isle of Wight.

Following transcription and analysis of both groups, the following themes were identified:

1. Purpose and function of the conceptual framework
2. Layout and design of the conceptual framework
3. Content of the conceptual framework: domain, sub-domain level and item level

The results are reported under these three themes for both groups separately.

5.3 Focus group 1

5.3.1 Purpose and function of the conceptual framework

The participants asked general questions about how the conceptual framework was going to be used and what scope it might have for use within practice. All participants reported that the definitions provided were important to the framework. It was suggested by one participant and agreed by all that an introduction to explain the purpose of the conceptual framework and guidance on how to look at the content would be useful.

5.3.2 Layout and design

Participants discussed how the iterative and changing nature of health could best be depicted. Everyone wanted to see the relationship between the central concept of 'health and wellbeing' to the four domains of physical functioning, mental wellbeing, social wellbeing and spirituality represented by double-headed rather than unidirectional arrows. Two participants suggested employing a graphic designer to assist with the design and layout once the content was agreed.

5.3.3 Content: domain, sub-domain level and component level

The relationships between each domain, sub-domain and its components were considered throughout the duration of the discussion. Practitioners when discussing how each domain related to the next and how health status was not static but a fluid process, suggested that further explanation would be required in any supporting documentation linked to the conceptual framework.

Spiritual wellbeing: there was some debate about the inclusion of this domain. One participant was concerned that the CST profession could be *"shooting themselves in the foot"* by including it, concerning the medical profession's stance. The other participants disagreed and felt that it was important to challenge medical views as spiritual wellbeing is important to clients.

Social wellbeing: One participant recommended that the domain of social wellbeing be developed to include: significant relationships, support networks, patient safety, dependants, health of partner and financial status**.** The other participants agreed**.**

Physical functioning: The domain of physical functioning, was discussed at component level in the context of 'working and earning a living' as part of 'everyday life', introducing the topics of 'digestion' and the 'patient's environment' as components were considered then rejected. It was suggested that 'energy levels' influenced all areas of health and wellbeing and should not, therefore, be depicted under just one domain.

Symptoms: the challenges of measuring symptoms were discussed. Pain, for example, was discussed in the context of the dimensions of severity, frequency and nature, noting that it could manifest as physical, spiritual, psychological or emotional pain. Participants suggested that the subjectivity of an individual's experience makes measurement difficult and the way that individuals learn to cope also has implications for measurement *"if somebody's ability to cope with pain has changed so much through their experience of the [CST] work, that their self-perception has changed so, I mean it's difficult"*. Pain was a motivator to get people to attend CST sessions.

Mental wellbeing: The meaning of the sub-domain termed 'self-efficacy/self-agency/self-care' was not clear to participants. Participants did not like the terms self-efficacy or self-agency. All participants emphasised that the terminology needed to be kept simple. Taking responsibility for one's health was deemed as important, as was its connection to self-care. Participants related to 'taking responsibility for one's health' from the perspective of clients making choices about their health and as part of a mental process that happens through a shift in awareness due to new insights about health. Participant 004 said *"it's probably the education that goes along with the treatment"*. The term 'being heart centred' was controversial and alternatives were suggested.

Other topics of discussion: The mental wellbeing sub-domain of 'emotions' and the spirituality sub domain of 'being present' were debated in relation to their position on the conceptual framework, but not what they represented.

Whether to include 'sexuality' on the conceptual framework and questionnaire were deliberated, 'intimate relationships' was suggested as a suitable term at conceptual framework level. Participants reported that 'animals' were important in people's lives and could impact health, yet had been missed from the conceptual framework.

General comments were made about the language and semantics, alternative words heard within practice were put forward to ensure acceptability to clients, for

example: confidence, empowered, focussed, resilience and connectedness. The overarching theme was considered satisfactory and central to the conceptual framework; participants wanted this explained in the conceptual framework's introduction.

5.3.4 Summary of focus group 1

The feedback from focus group 1 was practical and improved the clarity of the conceptual framework in relation to the design and layout. The suggestion of creating an introduction or a 'how to read the conceptual framework guide' and recruitment of a graphic designer was most useful in the early development phase. At this stage, between focus groups 1 and 2 'energy' was re-categorized from a component within the 'physical wellbeing' domain to a sub-domain located outside of the other domains, as 'energy' underpins all aspects of the conceptual framework. No further changes were made at this point but the comments relating to recommendations for additions to the social wellbeing domain, the inclusion of the spirituality domain, terminology of the mental wellbeing sub-domain of 'self-care, self-efficacy and self-agency', whether to include 'sexuality', and whether to re-categorize 'symptoms' were taken forward for use at the next focus group.

5.4 Focus group 2

5.4.1 Purpose and function of the conceptual framework

Participants assumed that the conceptual framework would be used within practice to provide an overview of the potential effects of CST, pinpointing those areas of a client's life where they might experience change. Participant 005 "*this is something I would like to use in my practice when I'm trying to explain what cranio-sacral therapy…. this feels really lovely to give the possibilities…"*.

5.4.2 Layout and design

One participant felt there was a subtlety in the depiction of the conceptual framework and liked the way it had been structured but the purpose of the heavy line placed around the central construct of health and wellbeing was unclear (see

Figure 7). Participants recognised that domains inherently overlap. One participant wanted scope for clients to personalise the framework to add their deficits.

5.4.3 Content: domains, sub-domains, items

Spiritual wellbeing: participants wanted this domain to include components that addressed a more philosophical element of spirituality and how individuals may integrate spirituality into their lives. One participant considered if an 'existential reference to existential being' was more relevant. Suggested components included: 'reflecting on the meaning of life (existential)' and 'living my core values'.

Social wellbeing: participants identified that 'engaging in life' had internal and external implications. The comments from Focus Group 1 were shared and discussed, the participants of Focus Group 2 disagreed with the previous group's thoughts about the inclusion of 'patient safety, dependants, health of partner and financial status' in to this domain. Yet, they liked the suggestion of 'significant relationships and support networks' being included. Recommendations were made to develop this domain to include 'engaging with local community and society'.

Mental wellbeing: Participants homed in on the topic of self-agency and self-care and, once again, this was most debated. Participant 005 emphasised their perspective about self-care, "*perhaps self-care is the most significant thing in wellbeing, is [having] the choices, the capacity and the motivation to take responsibility for one's own being".* The topics of 'sexuality' and 'financial security' were introduced for discussion based on the data in Focus Group 1. Two participants had concerns about including 'sexuality' on the conceptual framework, the other participant disagreed with them. Regarding 'financial security' one participant suggested that the topic of 'financial security' may influence a client's sense of self-worth and impact on health, the other practitioners disagreed, as the following quote explains. Participant 006 said "*I feel like what you've got here [items on the conceptual framework] is qualities about people and how they [are] relating to aspects of their lives, whereas, trauma, sexuality and money [are events or determinants that impact on health] more like facts and that feels different*

[rather than outcomes or health states]". The participants did not reach a consensus about either of these topics.

Physical functioning: It was proposed by NB that the sub-domain of 'symptoms' and its components be repositioned to outside of the other domains to reinforce the point that symptoms can manifest in any of the domains, participants agreed that this would be a good idea.

During reflections on this domain and its components, the sub-domain of 'sleep' was reframed based on discussions with supervisors and renamed 'sleep quality' and the component of 'disturbed' was deleted.

5.4.4 Summary of focus group 2

Participants discussed whether they might use the conceptual framework in practice with new clients as a tool to show the potential areas of effect that CST may have. There were no negative comments about the layout or design. Participants recognised that domains inherently overlap. Suggestions on how to improve the content of each domain were made. Participants liked the potential and scope the conceptual framework had for further research into the use of CST.

5.5 Revisions to the conceptual framework after focus groups 1 and 2

The conceptual framework was refined based on the discussions in the two focus groups.

Table 16: Revisions made to working conceptual framework after focus groups 1 and 2

Date	Revisions made to working conceptual framework of CST outcomes
July 2015	Graphic Designer assigned to enhance layout and design objective to make conceptual framework key identifiable in black and white.
July 2015	Previous qualitative work (Brough *et al*, 2015) had suggested that 'fostering awareness will change the way individuals will relate to the different aspects of self (mind, body, spirit), others and their environment', this statement was positioned at the top of the conceptual framework as a sub-heading. **Awareness** was depicted as a shaded area around the central concept of health and wellbeing, connected to the four domains of spiritual WB, Physical functioning, Mental WB, Social WB.
July 2015	**Domains** **Spiritual WB:** expanded to include components; 'living my core values', 'reflecting on the meaning of life (existential)'. **Physical Functioning:** 'symptoms' were repositioned on the conceptual framework to show that symptoms may not be just physical but can arise in other domains too. The component linked to 'sleep' was changed to 'sleep quality' and the component of 'disturbed' was deleted; 'work life' was linked to 'daily activities'; the component of 'energy level' was unlinked from this domain to show the wider implication 'energy levels' have on other domains. **Social WB:** this domain was expanded to include the components; 'making use of a support network'; 'feeling connected to local community' and 'intimate relationships'. **Mental WB:** the sub-domain of self-agency/self-efficacy/self-care was renamed **'inner involvement'**.

5.6 Focus group 3 – CST users

5.6.1 Demographics of focus group 3

Three female participants took part in the CST users focus group. Participant C001 had been having sessions on and off for about seven years, participant C002 had

been having sessions for a year and participant C003 had been having sessions for almost four years. All participants lived in Scotland and the focus group took place at the home of a CST practitioner in Edinburgh. No participants were known to NB or the host practitioner. The host practitioner acted as volunteer research assistant and ensured the recording equipment was working and made notes of the discussion for cross-referencing purposes. At no time was the research assistant involved in the discussions. The focus group discussion lasted for 1 hour 27 minutes.

5.6.2 Layout and design of conceptual framework

Layout: C001 liked the complexity of the conceptual framework *"I like the fact that it has a very definite centre to it, I mean in a way it's a target, your aim is health and wellbeing and all these other things are clearly related to it"*. A suggestion was made to ensure that the arrows on the conceptual framework were double headed to emphasise the nature of the relationships between all elements of the conceptual framework and that they are not exclusive. Some suggestions were made regarding consistency when applying grammar to the content.

5.6.3 Content of conceptual framework: domain, sub-domain level and item level

Spiritual wellbeing: One participant didn't like the word 'spirituality', and felt it could mean lots of different things. A second participant could see why it was on the conceptual framework and acknowledged that having CST can take one into deeper states, to what might be called 'spiritual experiences' which bring deeper philosophical insights when having sessions. Participant C002 agreed but felt the term 'spirituality' was too narrow to describe those experiences. In the end, all participants agreed that this domain should remain on the conceptual framework. The topic of 'animals' was talked about, in the context of being connected to something beyond oneself, nature and animals.

Physical Functioning: Participants discussed the components of the domain and queried whether it was intentional that 'daily activities' was only expressed as

'work life'. NB clarified that 'work life' was introduced due to the previous focus group discussions. Participants explored whether an individual's work life had consequences in other areas of life in addition to physical functioning. The group deliberated whether other aspects of life needed to be mentioned such as 'family life' which, to a mother of young children may be all encompassing. One participant questioned whether 'work life' was the correct term as it might exclude all other activities and suggested 'daily tasks' as an alternative. Participant C003 suggested that demands might be more suited *"it's just demands that are put on us"*.

Mental wellbeing: 'Inner involvement' was discussed, one participant thought that 'inner involvement' was a difficult concept to understand. The group were not sure that the term 'self-care' (which was one of the three the original terms used here) was correct either. One participant suggested that what they were discussing was 'taking responsibility for one's health' as it meant being pro-active in some way, *"it's more inner reflective where you've got to be more determined"*. C003 suggested *"taking time and responsibility for our own health"*. Two participants felt that taking reflective down time and allowing a nurturing experience were two separate things, participant C001 agreed.

Social wellbeing: participants were probed about 'intimate relationships' and what they thought about it. Participants had not had direct experience of CST influencing this aspect of their lives but all could see that it might be relevant for others and should, therefore, be on the conceptual framework.

Symptoms: participants queried the position of 'symptoms' as initially they expected to see it linked to physical wellbeing. NB suggested that 'symptoms' can be present in other areas such as emotional and psychological issues, all participants understood and agreed with its location on the conceptual framework.

Seen as 'missing' by CST users: Participants provided some 'missing items' from the conceptual framework:

- feeling connected to family history and past

- mindfulness
- autonomy
- relationships with animals

"Feelings of 'connectedness'" was discussed as an outcome because of having had CST, which was linked to 'trust' as 'trusting' was a key part of the process of CST. Participants suggested that 'connectedness' and 'trust' were linked to awareness and needed to be depicted in the centre of the conceptual framework with 'awareness'. Participants were keen that the following topics were documented as missing: healing (missing from the physical functioning domain), the CST process, and the value of the therapeutic relationship with a CST practitioner.

5.6.4 Summary of focus group 3

The domain of 'spirituality' was debated in terms of what it meant. 'Physical functioning' and the components linked to it were discussed and participants recommended 'daily tasks' replace 'work life'. At a sub-domain level, 'inner involvement' was examined in terms of what participants understood this to mean, and participants suggested that 'taking responsibility for self' was more appropriate. The position of 'symptoms' on the conceptual framework was queried until a consensus was reached that its current position was satisfactory. At a component level, participants agreed that 'intimate relationships' and the topic of 'animals' remain on the conceptual framework. Participants identified topics that they felt were missing from the conceptual framework and some recommendations to improve the layout and suggestions of grammatical consistency were made.

5.7 Revisions to conceptual framework based on CST users' perspectives

All of the topics discussed were given consideration before revisions were made to the working conceptual framework of CST outcomes. Table 17 reports the revisions after the third focus group.

Table 17: Revisions to conceptual framework of CST outcomes after focus group with CST users

Date	Revisions made to conceptual framework of CST outcomes
18/07/15	**Domains:** **Spiritual WB**: 'animals' were added to the component 'connection to self, divine, nature & ***animals***' **Physical Functioning:** the sub-domain of 'daily activities' was changed to '**daily tasks'**. **Mental WB:** linked to the sub-domain of 'self-concept', a component '**connection to family history and past'** was introduced. 'Heart centred' was deleted from the component 'sense of wholeness, self-confidence and replaced with **'self-concept'**. Inner involvement was altered to '**responsibility for self'**, the components linked to this sub-domain were separated in to two 'allowing a nurturing experience' and 'taking reflective down time', a component of **'autonomy'** was introduced. **Arrows:** the arrows linking symptoms and health and wellbeing were deleted, as was the arrow linked to energy. Arrowheads were assigned to both ends of all arrows to depict the multi-directional possibilities of change. **Layout:** A shaded circle was inserted around the central concept of health and wellbeing to highlight the term **'awareness'**.

Some topics and suggestions made in the three focus groups were not acted on:

- healing
- CST process
- therapeutic relationship with a CST practitioner

These topics have great relevance to CST, yet in this context were not outcomes but possibly mechanisms of action and are, therefore, not reported here. Some ideas were not taken forward as no agreement between participants could be reached (financial security and sexuality).

Figure 8, below, shows the revised conceptual framework of CST outcomes and Table 17, overleaf, lists the content of the conceptual framework inclusive of all changes made after the focus groups

Figure 8: Revised version of the conceptual framework of CST outcomes

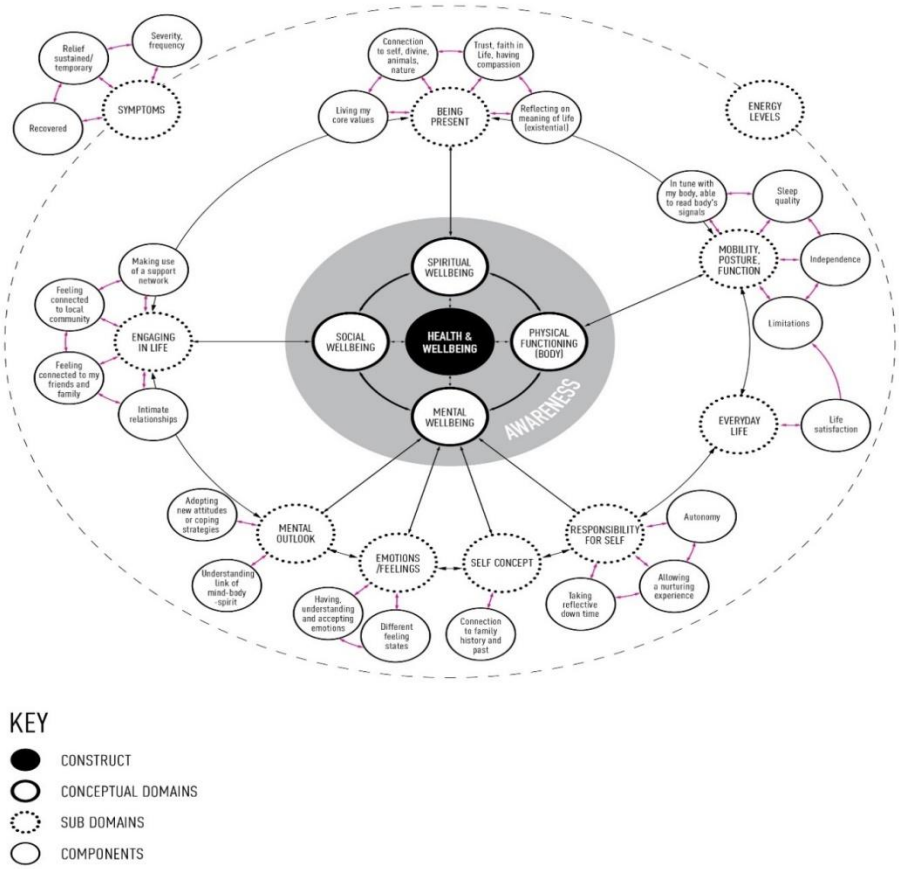


Table 18: Working conceptual framework of CST outcomes after revisions

Concept: Health and wellbeing		
Domain	**Sub-domain**	**Components**
Mental wellbeing	Responsibility for Self	Autonomy Allowing a nurturing experience Taking reflective downtime Sense of wholeness, self-confidence and balance
	Self-concept	Connection to family history and past
	Emotions and feelings	Different feelings state
	Mental outlook	Understanding and accepting emotions Understanding Mind-body-spirit links Adopting new attitudes or coping strategies
Physical wellbeing	Mobility, posture, function	In tune with one's body, able to read body's signals Sleep quality Independence Limitations
	Daily tasks	Life satisfaction Doing things I enjoy
Spiritual wellbeing	Being present	Reflecting on meaning of life Connection to self, divine, others, nature and animals Living my core values Trust, faith in life, having compassion
Social wellbeing	Engaging in life	Intimate relationships Making use of a support network Feeling connected to family and friends Feeling connected to local community
Symptoms (can manifest in any domain)		Severity, frequency Recovered Relief, temporary and sustained
Energy (is required for change in all domains and its components)		

5.8 Chapter summary

This section presented the working conceptual framework of CST outcomes and the way it was developed following focus group discussions. The domains of mental wellbeing, physical wellbeing, social wellbeing and spiritual wellbeing and the relevant components make up those domains and help to build a map of the ideas associated with this concept. This section then reported on the group participants' evaluation of the conceptual framework contents and layout and the revisions that were made as a result. Eight new components (living my core values, reflecting on the meaning of life, making use of a support network, feeling connected to the local community, intimate relationships, animals and autonomy, connection to family history and past) were added. Two were deleted (heart centred and disturbed sleep) and three items (daily tasks, sleep quality, responsibility for self) were renamed or reframed. Theorised links between sub-domains (energy, symptoms) and the other domains on the conceptual framework were also developed. The current version of the conceptual framework incorporating the revisions was presented in Figure 8.

Recommendations about how the conceptual framework might be used within practice were put forward by practitioners who took part in the evaluation, suggesting that the conceptual framework become a 'tool' in and of itself for CST practitioners to use within clinical practice to show the possibilities CST has in supporting the health and wellbeing of CST users.

In relation to this study, all the outcomes in the conceptual framework will need to be captured by a patient-reported outcome measure and, in part, the conceptual framework will be the foundation on which to appraise the content and face validity of candidate questionnaires.

6 Pre-testing preliminary WHHQ

A pool of 73 statements was generated based on previous qualitative work, exploring clients' experiences of CST (Brough *et al.*, 2015) and *via* discussions with experts. In this section, the results from the interviews carried out to pre-test the 73-item WHHQ are reported on. This process involved the clients' understanding of each statement, their evaluation of the content, layout, instructions, recall period and response options. Different terms for the middle point (sometimes, don't know, does not apply) were tested to ensure that the middle amount of the attribute was reflected and not the responders' inability to answer the question.

6.1.1 Demographics of the sample – semi structured interviews, rounds 1 and 2

Six semi-structured interviews (two rounds of three) were undertaken during the month of September 2015. Round 1 took place during 17th and 18th September and Round 2 on 28th September. Five female and one male participant represented each gender. Participants had been having CST sessions for as little as three months and up to seven years. The interviews took place in Edinburgh, Newcastle-upon-Tyne and Southampton. Interviews lasted a minimum of 42 minutes and maximum of 1 hour and 5 minutes.

6.2 Consensus meeting with CST practitioners

A consensus meeting with 16 CST practitioners from throughout the UK took place in Chedworth, Gloucestershire on 3rd October 2015, fifteen females and one male took part. All practitioners were trained in both CST and physiotherapy.

6.3 Design, layout and instructions

Round 1: Participants reported that they thought the layout of the questionnaire was clear, yet recommended that the headings of the different sections be shown more clearly. The two-tone differentiation between items was reported as confusing. All participants felt that the instructions were easy to follow. After round 1 of the interviews, changes to the design and layout including the removal of the two-tone effect and headings were set with a space above and below for

clarity. A question prompting participants to select the reasons for attending their session was added. Options included: physical functioning, emotional wellbeing, mental wellbeing, spiritual wellbeing and social wellbeing.

Round 2: Participants made no negative comments about the design, layout or instructions. All reported that the design, layout and instructions were satisfactory, clear and that they could complete the questionnaire without help.

Consensus Meeting: Overall, the practitioners reported that the design, layout and instructions were satisfactory. One group recommended alternative wording for the instructions about why responders are attending sessions. Revisions were made to reflect these recommendations. Revised instructions: 'Please tick the reasons you have come for your session (you can tick more than one box)'.

6.4 Recall period – 2 weeks

Round 1: Most but not all participants thought the two-week time frame was satisfactory; one preferred a one-week time frame considering possible inability to recall.

Round 2: The participants in this round were satisfied with the recall period; participant C005 felt it may depend on how often someone was having sessions, but felt confident that a recall period of two weeks was satisfactory, and said *"generally we can probably remember how we have felt in the past two weeks"* (pg. 2). C007 believed in their experience, CST had an immediate effect and that a weekly recall period may also be suitable. No changes were made to the recall period based on the patient reports in round 2.

Consensus meeting: There was agreement amongst the groups that the two-week recall period was acceptable.

6.5 Name of the questionnaire

Round 1: Participants asked why CST was not in the title, but on informing them that the questionnaire may have wider appeal, they raised no concerns. One participant claimed that she hadn't noticed the title *"I didn't really notice the title, if*

you hadn't asked the question I wouldn't have noticed the title" pg. 3 (C002). Two participants asked why 'Warwick' was in the title, on explaining that Warwick was the University linked to this study, they had the clarity they needed.

Round 2: Two participants questioned if the term 'holistic' was widely understood. Everyone in this round expected to see 'CST' in the questionnaire name. Participants were told that the questionnaire may have wider application and they understood why it had been left out and a more generic name given. No changes were made to the name of the questionnaire after round 2.

Consensus meeting: The groups had different opinions about the suggested name; one group suggested changing 'holistic' for 'wellbeing'; a second group wanted 'CST' in the title; the third group felt that the current title was 'a mouthful' and that an acronym might be more appropriate; the fourth group felt the questionnaire name was acceptable. The questionnaire name was not changed based on the CST practitioner discussions.

6.6 Response options

Round 1: Two different sets of response options were presented to the interview participants based on the syntax of each statement. Set one, *'none of the time, rarely, sometimes, often and all the time'* and set two, '*disagree strongly, disagree somewhat, don't know, agree somewhat and agree strongly'*. The options were discussed in relation to each question and timeframe. Two participants felt that some statements were more philosophically orientated and, therefore, needed to be posed based on context and would require a longer time frame during a course of treatments not within one session. The response options were not changed after round 1.

Round 2: One participant felt that having the combination of two sets of response options on the questionnaire was confusing. All participants referred to the response options whether, in their opinion, the item in question could be given a clearer response. For example: Item 17, 'I've been able to adapt to meet life's challenges' participant C005 wondered if it was not relevant to a responder

whether a 'not applicable' option could be available? Based on the feedback from participants it was decided that the response options were better not including a neutral category such as 'don't know'. At this point, items were revised to ensure that the adjectives were appropriate for the stem and endorsed response options.

Consensus meeting: The groups reported in the context of the items they were assigned. In general, both sets of response options were acceptable, yet one group recommended an additional option if the responder was unable to answer the statement, for example, not applicable or not sure. A second group were concerned that changes between responses on the questionnaire could be confusing. No changes were made to the response options but instructions were added to the questionnaire to inform when the response options were about to change. Based on the reports of the interviews and consensus meeting, thirty-six statements on the 52-item WHHQ had the first set of response options (set one) and sixteen statements (17,19,20,22,23,24,25,27,28,30,31,32,33,35, 66,67) had responses from set two.

6.7 Content

The 73 statements were presented in groups related to each domain; (the number of statements in each group is shown in brackets) physical functioning (5), symptoms (4), everyday life (6), mental wellbeing/mental outlook (8), self-concept (4), responsibility for self (8), emotions and feelings (6), spirituality (9), social wellbeing (6), holistic wellbeing (17). Table 19 shows the 73 statements on the original list as presented to participants and the results of each round of interviews. Items were removed if they were found to be unclear, difficult to understand or confusing, or duplicated. Items were revised to improve understanding or to ensure tense consistency. Round 1, 17 items were removed and 1 item was revised. In round 2, 3 items were removed and 26 items were revised. Based on the consensus meeting reports, 3 items were removed and 10 items were revised. Two rounds of cognitive interviews and the consensus meeting resulted in the removal and rewording of items leaving a 52-item version of the WHHQ.

Table 19: Summary of item changes per interview rounds based on patient reports and consensus meeting with CST practitioners

Item at pre-test	**Round 1**	**Round 2**	**Consensus meeting**
Physical functioning			
1) I've been physically independent	Revised – I've been physically well	No change	No change
2) I've been able to read my body's signals	No change	No change	Revised – I'm aware of my body's needs
3) I've had lots of energy	No change	No change	No change
4) I've been sleeping well	No change	No change	No change
5) I've been in tune with my body	Removed		
Symptoms			
6) I've been in pain	No change	No change	No change
7) My symptoms have been a problem	No change	No change	Removed
8) I've had relief from my symptoms (reverse code)	No change	No change	No change
9) My health limits my daily activities	No change	Revised – my symptoms limit my daily activities	No change
Everyday life			
10) I feel satisfied by my school, work or current role in life	No change	Revised – I've felt satisfied…	No change
11) I feel satisfied with my work/life balance	No change	Revised – I've felt satisfied…	No change
12) I feel satisfied with my life overall	Removed	Removed	
13) I have had too many demands made on me	No change	No change	No change (reverse code)
14) I've engaged in nurturing activities	No change	Revised - I've nurtured myself	Revised – I've looked after my own needs

15) My daily life is full of things that keep me interested	No change	No change	No change

Item at pre-test	Round 1	Round 2	Consensus meeting
Mental wellbeing/mental outlook			
16) I live life one day at a time and don't really think about the future	Removed		
17) I've been able to adapt to meet life's challenges	No change	No change	No change
18) Life has been a continuous process of learning, changing and growth	No change	Removed	
19) I've been feeling positive about life	No change	No change	No change
20) I've been coping with daily life	No change	No change	No change
21) I believe my health is linked to my daily life	Removed		
22) The perspective I have about my condition helps me manage my life	No change	Revised – my awareness about my health helps me manage my life	No change
23) I tend to bounce back quickly after hard times	No change	Revised – I've felt resilient	No change

Item at pre-test	Round 1	Round 2	Consensus meeting
Self-concept			
24) I keep on learning about myself and my body	No change	Revised – I'm learning about myself and my body	No change
25) I judge myself by what I think is important, not by the values of what others think is important	No change	Revised – I view myself by what I think is important, not by the values of others	No change
26) I've felt in balance	No change	No change	Removed
27) I've been feeling confident	No change	No change	No change
Responsibility for self			
28) Taking an active role in my own health is the most important factor in determining my wellbeing	No change	Revised – I'm in charge of my health and wellbeing	Revised – I'm in control of my health and wellbeing
29) I deal consciously with myself	Removed		
30) I feel able to take care of myself	No change	Revised – I've felt able to take…	No change
31) I've taken reflective downtime	No change	Revised – I've been able to stop and reflect	No change
32) I've asked for help when I've needed it	No change	No change	No change
33) Making sense of the past helps me figure out what to do in the present	No change	No change	Revised – I've been troubled by issues from my past
34) I feel able to take care of myself	Removed – duplicate		
35) I've identified activities that I enjoy	No change	No change	Revised – I've taken time to do things I enjoy

Item at pre-test	Round 1	Round 2	Consensus meeting
Emotions and Feelings			
36) I've been feeling anxious (reverse code)	No change	Revised – I've felt anxious	No change
37) I've been feeling confident	Removed – duplicate		
38) I've been able to express how I feel	No change	No change	No change
39) I've been feeling calm	No change	Revised – I've felt calm	No change
40) I've been feeling joyful	No change	Revised – I've felt joyful	No change
41) I've been feeling sad (reverse code)	No change	Revised – I've felt sad	No change
Spirituality			
42) I've felt my inner strength	No change	No change	No change
43) I've trusted others	No change	No change	No change
44) I've been compassionate to myself	No change	No change	No change
45) I've been compassionate towards others	No change	No change	No change
46) I've had faith in life's journey	Removed		
47) I reflect on the meaning of life	No change	Revised – I've reflected on the meaning of life	Revised – I've felt my life has meaning
48) I've been feeling connected to God or the divine	Removed		
49) I've felt connected to nature	No change	No change	No change
50) I have a spiritual orientation in life	No change	Revised – I've felt spiritually orientated	Removed

Item at pre-test	Round 1	Round 2	Consensus meeting
Social wellbeing			
51) I interact consciously with others	Removed		
52) I interact consciously with my environment	Removed		
53) I've been feeling engaged in life	No change	Revised – I've felt engaged in life	No change
54) I've felt connected to my friends and family	No change	No change	No change
55) I've identified people who help me take care of myself	No change	No change	Revised – I have people in my life who help me take care of myself
56) I have some meaningful relationships	Removed		
Holistic wellbeing			
57) I've had a sense of wellbeing	No change	No change	No change
58) I've been aware of the connection between mind, body and spirit	No change	No change	No change
59) I trust my intuition	No change	Revised – I've trusted my gut instinct	No change
60) I'm self-aware	No change	No change	Removed
61) I strive for more broad awareness	Removed		
62) I'm living authentically	No change	Revised – I've felt like a fraud (reverse code)	No change
63) I've been feeling well	No change	Revised – I've felt well	No change
64) I feel more connected to my body	Removed		
	Removed		

65) I feel more aware of the way my mind and body work together	No change	No change	Removed
66) I feel more aware of the way my mind and body are affected by my physical environment	No change	No change	Revised – I'm aware that my mind and body are affected by my physical environment
67) I feel more aware of the way my mind and body are affected by my social environment	No change	No change	Revised – I'm aware that my mind and body are affected by my social environment
68) I am more aware of what my body needs to stay healthy	No change	Revised – I've been aware of what my body….	No change
69) I can read my body's signals better	No change	Revised – I've read my body signals well	No change
70) I can trust my body to tell me what it needs	Removed		
71) I feel less rejected of my body and its symptoms (reverse code)	No change	Revised – I've felt rejecting of my body and its symptoms	Revised – I've found it hard to accept my body and its symptoms
72) I feel less ashamed of the way my body is currently working	No change	Revised – I've felt ashamed of my body	No change
73) I feel less angry at my situation now	No change	Revised – I've felt angry	No change
			End.

6.8 Pre-testing of 52-item WHHQ - cognitive interviews, round 3

The 52-item WHHQ was pretested during a round of cognitive interviews in October 2015. This section reports on the feedback from participants in assessing the content validity.

6.9 Demographics of the sample, round 3

Three participants took part in the cognitive interviews, all were female. Participant C008 was 34 years of age and had 30 sessions of CST fortnightly, participant C009 was also 34 years of age and had been having monthly sessions of CST for five years, participant C010 was 38 years of age and had received five CST sessions at the time the interviews were undertaken. Interviews were held in Bristol and London.

6.10 Interviews

Each interview lasted on average 25 minutes. The mean time it took participants to complete the questionnaire was 7.7 minutes (4.5 minutes, 7.5 minutes and 11 minutes).

6.11 Design, layout, instructions

Participant C010 suggested that the instructions for completing the questionnaire be on the top of every page. C010 reported that the headings included in the questionnaire were useful at first glance, but once she had engaged in the process of answering each item they became less significant, yet commented that the order in which the items were presented assisted in her thought process. References were made to the motivation for having CST (a bad neck, physical pain) and how that may then influence emotions and mental wellbeing, as she expected to see mental wellbeing after the physical functioning statements "*it's nice to notice all the different facets that are affected, that the treatment can help…it helps clarify the direction for it [the journey through the questionnaire]"* pg. 6, (C010). No further recommendations were made to the design, layout and instructions.

6.12 Response options, round 3

Two participants found item 33 'I've been troubled by issues from my past' difficult to answer with the response options from set 2 ('*disagree strongly, disagree somewhat, don't know, agree somewhat and agree strongly*'). Participant C009 had responded mainly with the option of 'often', when prompted to explain this, it was reported that 'all of the time' *"is quite extreme" pg.2* and questioned if there might be an alternative *"maybe there is something between 'often' and 'all the time'"* pg. 2, (C009). A second participant agreed that 'all of the time' was extreme. The 'majority of the time' was put forward and the participants concurred that 'majority of the time' would be a suitable addition to 'all'. One participant commented on how reverse coded items meant that an adjustment was required in their orientation to the statements *"there's a swap around from a positive to that [negative item], it took my brain a little while to engage as to what it meant"* pg. 2 (C008). Based on these reports the response option of 'all of the time' was revised to 'all or most of the time'.

6.13 Content, round 3

Statements which appeared to slow the responders down and required *"more thinking time"* pg. 9, (C010) were those longer in length, for example: 25, 'I view myself by what I think is important, not by the values of others', 33 'I've been troubled by issues from my past', 66 'I'm aware that my mind and body are affected by my physical environment' and 67 'I'm aware that my mind and body are affected by my social environment'. One participant felt that the statement 10, 'I've felt satisfied by my school, work or current role in life' and 11, 'I've felt satisfied with my work/life balance' addressed the same point. No statements were removed or revised based on round 3 of the cognitive interviews.

6.14 Chapter summary

In summary, two rounds of semi-structured interviews and one round of cognitive interviews were undertaken during the months of September and October 2015 with 9 participants and a consensus meeting with 16 CST practitioners was also held in October 2015. A list of 73 items was evaluated for inclusion in the draft version

of the WHHQ. Twenty-one items were deleted during the evaluation, 36 items were revised to improve the comprehensibility of the statements (1 item was revised after round 1, 26 items were revised after round 2, 9 items were revised after the consensus meeting with CST practitioners). No changes were made to the content after the cognitive interviews, but the response options were revised to 'all or most of the time' based on CST user reports.

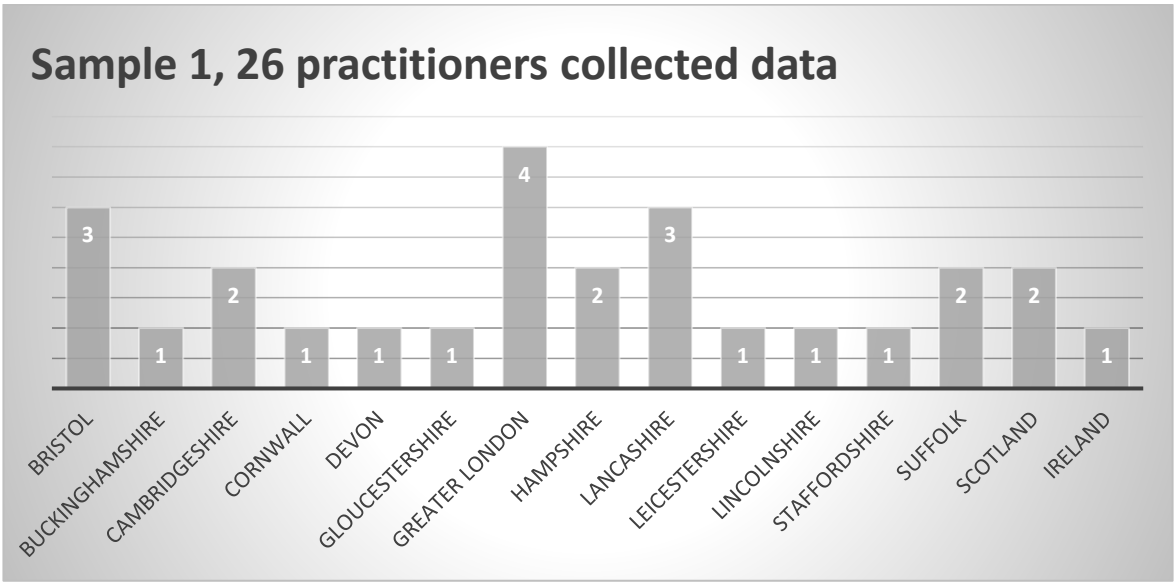
7 Psychometric testing of the 52-item WHHQ

The 52-item WHHQ was used with a sample of CST users to evaluate the measurement properties of the questionnaire.

7.1 CST practitioners who collected data

Fifty-five practitioners expressed an interest in collecting data; 8 withdrew, 21 did not respond, 26 participated. Four hundred and five 52-item WHHQ were sent to 26 practitioners. The counties in England where practitioners are located can be seen from Figure 9 below. Ireland and Scotland are also represented:

Figure 9: Number of practitioners and data collection locations



7.2 Demographics of sample 1

There were 142 participants in this study, a 35% response rate. A heterogeneous sample, made up of individuals with a variety of disease states and severities, was recruited, promoting a wide range of responses. The demographics of sample 1 are presented below.

7.3 Gender

Of the participants in this study, 81% were female and 19% male.

7.4 Age of participants

The participants in this study had an age range between 16 years and 83 years. The age categories include: teenagers (16-20), young adults (21 – 39), middle aged (40 – 64) and mature (65 +), with the highest proportion of the sample (44.4%) being 40-64 years and most of this sub-sample were women, as shown in Table 20.

Table 20: Age of participants by gender

		Female	Male	Total	Percent
Age	16-20 years	7	3	10	7
	21-39 years	26	3	29	20.4
	40-64 years	54	9	63	44.4
	65 + years	27	12	39	27.4
	Not given	1		1	0.7
	Total	115	27	142	

7.5 Number of CST sessions undertaken and year of first CST session

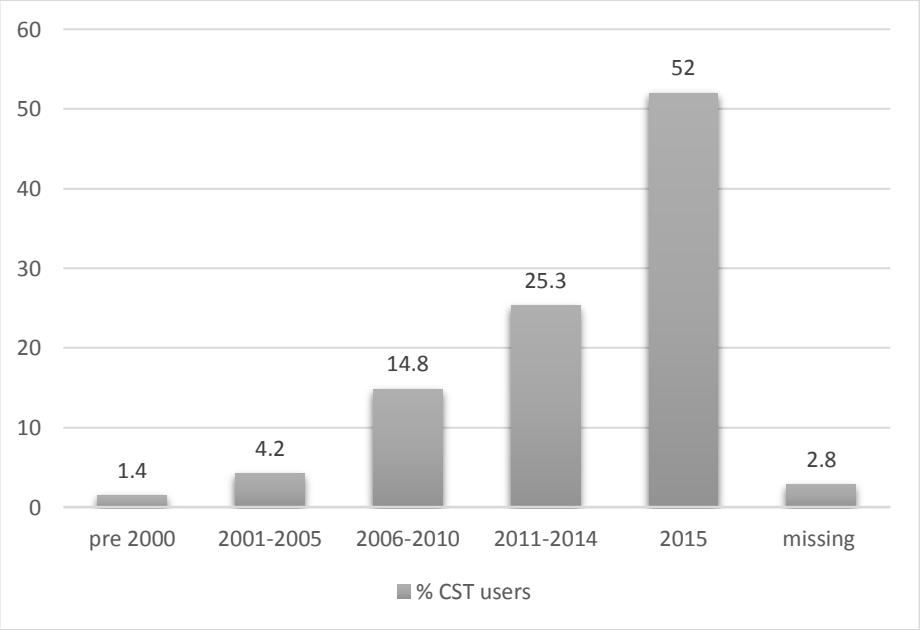
The number of CST sessions each participant had undertaken at the time they completed the 52 item-WHHQ are shown in Table 21. Here, 0 represents those participants undertaking their first session of CST. The gender split is also shown.

Table 21: Sample 1, number of CST previous sessions received by gender

		Gender		
		Female	Male	Total
Number of sessions	0	8	3	11
	1 - 12	48	8	56
	13 - 24	22	7	29
	25 - 52	21	5	26
	53+	12	4	16
	Missing data			4
	Total	111	27	142

Figure 10 shows the year in which CST users had their first CST session total n = 142, 52% of the sample started sessions during 2015, the year in which this study was carried out. 1.4% of the sample had been having CST sessions for more than fifteen years, 14.8% for almost ten years and 25.3% for up to four years.

Figure 10: Sample 1, year of first CST session



7.6 Self-reported overall wellbeing

When completing the 52 item-WHHQ participants were also asked an anchor question to rate their overall wellbeing. The item was "How would you rate your overall wellbeing today?"

Table 22: Sample 1, self-reported overall wellbeing by gender

		Gender		
		Female	Male	Total (%)
Self-rated Overall Wellbeing	Poor	2	2	4 (3%)
	Fair	32	7	39 (27%)
	Good	54	12	66 (47%)
	Very good	23	6	29 (21%)
	Excellent	2	0	2 (1%)
	Missing			2 (1%)
Total		113	27	142

Table 22 shows how participants responded to the anchor question: 47% of the participants rated their overall wellbeing as 'good'. With only 3% of the participants rating their health as 'poor'.

7.7 Reasons for having sessions

Participants were asked to report why they were coming for CST sessions see Table 23. Multiple reasons were possible and participants were prompted to select all the reasons that were relevant to them, from the following options: mental wellbeing, holistic wellbeing, social wellbeing, spiritual wellbeing, problems with emotions and problems with body.

Table 23: Sample 1 - Reasons for having CST sessions

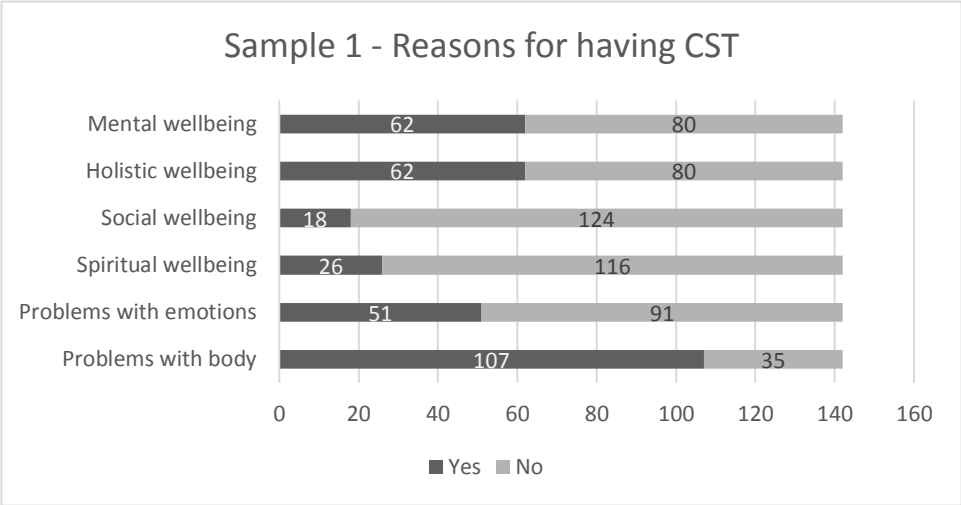


Table 24 shows the gender split of why participants were having sessions.

Table 24: Reasons for attending CST sessions by gender

Reason for sessions (n %)	Female (N=115)	Male (N=27)	All sample (N=142)
Problems with body	87 (76)	20 (74)	107 (75)
Problems with emotions	43 (37)	8 (30)	51 (36)
Spiritual wellbeing	23 (20)	3 (11)	26 (18)
Social wellbeing	16 (14)	2 (7)	18 (13)
Holistic wellbeing	52 (45)	10 (37)	62 (44)
Mental wellbeing	49 (43)	13 (48)	62 (44)

7.8 Construct validity

This section explores the factor structure 52-item WHHQ firstly, through hypothesized conceptual framework domains and, secondly, reporting the results of an exploratory factor analysis.

There were no clear-cut ideas about the number of dimensions in the proposed conceptual framework. It was hypothesized that the 52-item WHHQ represented six domains which were all correlated. Table on page 159 shows the six domains and the items hypothesised to represent each domain based on the conceptual framework of CST outcomes.

Table 25: 52-Item WHHQ hypothesized domains and representing statements

Domains	Item No	Statements
Spiritual wellbeing	42	I've felt my inner strength
	43	I've trusted others
	44	I've been compassionate to myself
	45	I've been compassionate towards others
	49	I've felt connected to nature
	47	I've felt my life has meaning
	53	I've felt engaged in life
	54	I've felt connected to my friends and family
	55	I have people in my life who help me take care of myself
	10	I've felt satisfied by my school, work or current role in life
	57	I've had a sense of well being
	59	I've felt satisfied with my work/life balance
	62R	I've trusted my gut instinct
	63	I've had too many demands made on me
	68	I've felt like a fraud
	69	I've looked after my own needs
	71R	I've felt well
	72R	My daily life is full of things that keep me interested
	58	I've been aware of what my body needs to stay healthy
	66	I've read my body's signals well
	67	I've felt it hard to accept my body and its symptoms
Mental wellbeing	17	I've felt ashamed of my body
	20	I've been aware of the connection between mind, body and spirit
	22	I'm aware that my mind and body are affected by my physical environment
	23	I'm aware that my mind and body are affected by my social environment
	19	I've been feeling positive about life
	24	I'm learning about myself and my body
	28	I'm in control of my health and wellbeing
	27	I've been feeling confident
	30	I've felt able to take care of myself
	31	I've been able to stop and reflect
	32	I've asked for help when I've needed it
	35	I've taken time to do things I enjoy
	25	I view myself by what I think is important, not by the values of others
	33R	I've been troubled by issues from my past

Domains	Item No	Statements
Physical Functioning	1	I've been physically well
Social wellbeing	2	I'm aware of my body's needs
	3	I've had lots of energy
	4	I've been sleeping well
Everyday Life (PF)	10	I've felt satisfied by my school, work or current role in life
Holistic wellbeing	11	I've felt satisfied with my work/life balance
	13R	I've had too many demands made on me
	14	I've looked after my own needs
	15	My daily life is full of things that keep me interested
Emotions and Feelings (MWB)	36R	I've felt anxious
	38	I've been able to express how I feel
	39	I've felt calm
	40	I've felt joyful
	41R	I've felt sad
	73R	I've felt angry
Symptoms	6	I've been in pain
	8R	I've had relief from my symptoms
	9	My symptoms limit my daily activities

Headings in italics represent sub-domains physical functioning and mental wellbeing.

7.9 Exploratory factor analysis results

Exploratory factor analysis was undertaken to investigate the factor structure of the WHHQ and as an item reduction technique.

7.10 Item reduction using exploratory factor analysis

As described in the methods section on page 117, the 52 items (Table 25) were reduced to a more clinically manageable instrument length. Initially, three Models were estimated with one item deleted on each occasion (see Table 26 for details). Kappa was set to 4 and the number of factors to extract was not specified. Model 4 failed to converge in 25 iterations, hence the EFA settings were adjusted: increasing the maximum iterations for convergence to 50, and reducing the Kappa to 2 (leading to lower correlations among factors and allowing for more complexity in the structure). A further 20 models were undertaken iteratively using these parameters, resulting in a total of 23 deleted items.

To refine the remaining 29 items further, the EFA parameters were changed to specify the number of factors to be extracted as 6, the total variance extracted is no less than 40% for all extracted factors. Another 2 Models (numbers 24 and 25) were estimated and an item deleted after each iteration. The EFA parameters were adjusted so that the items were extracted based on the Eigenvalues > 1 rule, for Models 26 – 31. To ensure the total variance extracted was no less than 40% for all extracted factors the EFA parameters were adjusted to specify the number of factors to be extracted as 4 (Models 32 – 34).

A total of 32 items were deleted in this process, Table 26 lists the order in which the redundant items were deleted and reasons for deletion.

Table 26: Item reduction list

Model number	statement	item number	Reason for deleting
Model 1	I've been sleeping well	4	triple loaded factors 6, 10, 11
Model 2	I've felt angry	73R	double loaded factors 5 and 10
Model 3	I've felt connected to my family and friends	54	double loaded factors 1 and 8
Model 4	**Rotation failed to converge in 25 iterations**	NA	NA
Model 5	I've been satisfied with my work/life balance	11	double loaded factors 1 and 2
Model 6	I've been feeling positive	19	double loaded factors 1 and 2
Model 7	I've asked for help when I've needed it	32	double loaded factors 2 and 11
Model 8	I've been aware of the connection between mind body and spirit	58	double loaded factors 4 and 7
Model 9	I've felt connected to nature	49	double loaded factors 1 and 8
Model 10	I've felt my inner strength	42	double loaded factors 4 and 9
Model 11	I've been able to adapt to meet life's challenges	17	double loaded factors 2 and 7
Model 12	I've been able to express how I feel	38	double loaded factors 1 and 11
Model 13	I've read my body's signals	69	double loaded factors 4 and 9
Model 14	I've looked after my own needs	14	double loaded factors 4 and 9
Model 15	I've had people in my life who help me take care of myself	55	double loaded factors 7 and 10
Model 16	I've felt resilient	23	double loaded factors 2 and 4
Model 17	I've had too many demands made on me	13R	Solitary item
Model 18	I'm aware that my mind and body are affected by my physical environment	66	Solitary item
Model 19	I'm aware that my mind and body are affected by my social environment	67	Solitary item

Model 20	I view myself by what is important to me not by the value of others	25	triple loaded factors 4, 5 and 6
Model 21	I've felt it hard to accept my body and its symptoms	71R	Solitary item
Model 22	I've felt well	63	triple loaded factors 3, 7 and 8
Model 23	I've had a sense of wellbeing	57	triple loaded factors 1, 4 and 7

Ran model using less factors to reduce variance. Extracted Factors 6 kappa 2 iterations 50			
Model 24	I've been compassionate towards others	45	triple loaded factors 1, 5 and 7
Model 25	I've trusted others	43	double loaded factors 1 and 6
EFA parameters adjusted so that items would be extracted based on Eigenvalues > 1 rule			
Model 26	I've been troubled by issues from my past	33R	double loaded factors 6 and 7
Model 27	I've felt sad	41R	double loaded factors 1 and 5
Model 28/34	I've been feeling confident	27 & 37	double loaded factors 2 and 4
Model 29	I've had relief from my symptoms	8R	double loaded factors 3 and 6
Model 30	I've been compassionate towards myself	44	double loaded factors 2 and 5
Model 31	I've felt anxious	36R	double loaded factor 1 and 5 weak communalities

Ran model using less factors to reduce variance. Extracted Factors 4 Kappa 2 iterations 50			
Model 32	I've been coping with daily life	20	double loaded factors 1 and 2 weak communalities
Model 33	I've taken time to do things I enjoy	35	double loaded factors 1 and 2
Model 34	No items met exclusion criteria	NA	NA

7.11 Final reduced item Model

The final reduced item Model was Model 34. The pattern matrix is shown in Table 28 and the total variance related to the four extracted factors is explained in Table 27.

Table 27: Total variance explained

Extraction Sums of Squared Loadings		
Total	% of Variance	Cumulative %
5.661	29.7	29.7
2.118	11.1	40.9
1.707	8.9	49.9
1.431	7.5	57.4

The Scree plot (Figure 11) using the Eigenvalue > 1 rule suggests a strong first factor and four, possibly, five extra factors.

Figure 11: Scree plot for 19 items (Model 33)

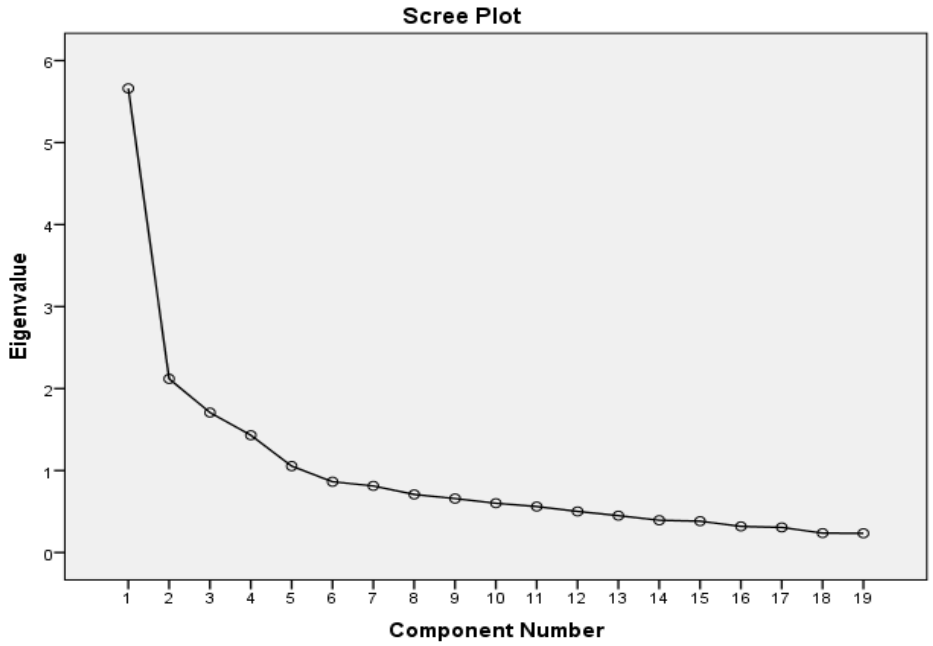


Table 28: Pattern Matrix of Model 34 – draft WHHQ 2: 19 items of 4 factors

			Component/Loadings*			
Order	**Item ref**	**Statement**	1	2	3	4
1	15	My daily life is full of things that keep me interested	0.923			
2	53	I've felt engaged in life	0.82			
3	47	I've felt my life has meaning	0.748			
4	10	I've been satisfied by my school, work or current role in life	0.743			
5	40	I've felt joyful	0.731			
6	39	I've felt calm	0.587			
7	22	My awareness about my health helps me manage life		0.794		
8	31	I've been able to stop and reflect		0.78		
9	24	I'm learning about myself and my body		0.757		
10	28	I'm in control of my health and well being		0.701		
11	2	I've felt aware of my body's needs		0.563		
12	30	I've felt able to take care of myself		0.549		
13	6R	I've been in pain			-0.849	
14	9R	My symptoms limit my daily activities			-0.849	
15	1	I've been physically well			0.754	
16	3	I've had lots of energy			0.559	
17	72R	I've felt ashamed of my body				0.841
18	62R	I've felt like a fraud				0.831
19	59	I've trusted my gut instinct				0.384

*Absolute loadings < 0.3 have been suppressed

7.12 Face and content validity cross check

The data gathered from CST users and practitioners during the qualitative methods identifying items of importance, were revisited to ensure that face and content validity was maintained. Table 29 below shows eleven items which all CST users reported as important, yet had not been retained during the exploratory factor analysis process.

Table 29: Items to be reconsidered for inclusion in the WHHQ

Order	Item ref	Item long name
1	4	I've been sleeping well
2	13R	I've had too many demands made on me
3	42	I've felt my inner strength
4	55	I've had people in my life who have helped me to take care of myself
5	57	I've had a sense of wellbeing
6	66	I'm aware that my mind and body are affected by my physical environment
7	67	I'm aware that my mind and body are affected by my social environment
8	32	I've asked for help when I've needed it
9	38	I've been able to express how I feel
10	36R	I've felt anxious
11	75	My sexual needs are being met

At a consensus meeting, CST practitioners were asked to re-evaluate each of the above items for inclusion in the WHHQ. They were asked to consider if:

- the omission of any of these items would prevent practitioners from using the questionnaire (poor face and content validity)?
- the addition of the item(s) would influence response rates?
- the items would be sensitive to change over time?
- the item would add any additional information to the scale?

This resulted in item reference numbers 36, 55, 57, 66 and 67 confirmed as deleted as the items were deemed not sensitive to change over time, or being too

subjective. Item 75 was deleted due to the lack of consensus amongst practitioners, as some felt having an item related to sexuality could negatively influence response rates. Hence the remaining six items were returned to the potential item pool for the WHHQ, making the 25-item WHHQ.

8 Structural validity

In this section the pattern matrix and component correlation matrix for the draft 25-item WHHQ are presented.

8.1 Pattern and correlation matrices

The data set response patterns of the items on the 25-item WHHQ were extracted and analysed. An oblique rotation (Promax) EFA was performed.

As an oblique rotation (Promax) allows the factors to correlate, the loadings and correlations are not distinct. Each row on the pattern matrix shows loadings as the standardized observed variables expressed as a function of the factors, see Table 30.

Table 30: Draft 3 WHHQ: Pattern Matrix 25-item WHHQ

Item long name	Component*			
	1	2	3	4
1. My daily life is full of things that keep me interested	.790			
2. I've felt engaged in life	.725			
3. I've felt my life has meaning	.689			
4. I've been satisfied by my school, work or current role in life	.652			
5. I've felt joyful	.653			
6. I've felt calm	.564			
7. My awareness about my health helps me manage life		.714		
8. I've been able to stop and reflect		.753		
9. I'm learning about myself and my body		.722		
10. I'm in control of my health and wellbeing		.685		
11. I'm aware of my body's needs		.459		
12. I've felt able to take care of myself		.560		
13. I've been in pain			-.784	
14. My symptoms limit my daily activities			-.838	
15. I've been physically well			.719	
16. I've had lots of energy			.579	
17. I've felt ashamed of my body				.781
18. I've felt like a fraud				.792
19. I've had too many demands made on me	.406			
20. I've trusted my gut instinct		.379		.350
21. I've been sleeping well	.373			
22. I've felt my inner strength		.465		
23. I've felt connected to my friends and family	.572			
24. I've asked for help when I've needed it		.537		
25. I've been able to express how I feel	.576			

*Absolute loadings <0.3 have been suppressed

The component correlation matrix displays the correlation coefficients between a single variable and every other variable in the overall structure. As seen in Table 31, the correlations are not high but the factors are all correlated.

Table 31: Component Correlation Matrix 25-item WHHQ

Component	1	2	3	4
1	1.000	.370	.193	.103
2	.370	1.000	.137	.096
3	.193	.137	1.000	.049
4	.103	.096	.049	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Promax with Kaiser Normalization.

8.2 25-item WHHQ

The 25 item WHHQ with the chosen response options for each item is shown in Table 32.

Table 32: 25-Item WHHQ with response options

		Never	Rarely	Sometimes	Often	Most or all the time
15	My daily life has been full of things that keep me interested	0	1	2	3	4
53	I've felt engaged in life	0	1	2	3	4
47	I've felt my life has meaning	0	1	2	3	4
10	I've felt satisfied by my school, work or current role in life	0	1	2	3	4
40	I've felt joyful	0	1	2	3	4
39	I've felt calm	0	1	2	3	4
22	My awareness about my health has helped me manage life	0	1	2	3	4
31	I've been able to stop and reflect	0	1	2	3	4
24	I'm learning about myself and my body	0	1	2	3	4
28	I've felt in control of my health and wellbeing	0	1	2	3	4
2	I've felt aware of my body's needs	0	1	2	3	4
30	I've felt able to take care of myself	0	1	2	3	4
6R	I've been in pain	4	3	2	1	0
9R	My symptoms have limited my daily activities	4	3	2	1	0
1	I've been physically well	0	1	2	3	4
3	I've had lots of energy	0	1	2	3	4
72R	I've felt ashamed of my body	4	3	2	1	0
62R	I've felt like a fraud	4	3	2	1	0
13R	I've had too many demands made on me	4	3	2	1	0
59	I've trusted my gut instinct	0	1	2	3	4
4	I've been sleeping well	0	1	2	3	4
42	I've felt my inner strength	0	1	2	3	4
54	I've felt connected to my family and friends	0	1	2	3	4
32	I've asked for help when I've needed it	0	1	2	3	4
38	I've been able to express how I feel	0	1	2	3	4
				(Maximum score 100)	Questionnaire Total	

9 Psychometric evaluation of the 25-item WHHQ

A new sample of CST users, who had not completed the 52-item WHHQ before, were asked to complete four questionnaires twice, once before a session of CST and once before their next session of CST.

Of the 66 practitioners who registered their initial interest in data collection for this sub-study: 25 (38%) did not respond, 15 (23%) withdrew, 26 (39%) participated.

Of the 26 practitioners who collected data: 24 (92%) were based in the UK, 1 (4%) was from Australia and 1 was (4%) from Italy (seen in Figure 12). All data were collected from CST users whose first language was English.

Figure 12: Location of UK and non-UK practitioners for sample 2

	Location	**No. Practitioners**
County of England	**Buckinghamshire**	1
	Cambridgeshire	1
	Devon	1
	Dorset	1
	Gloucestershire	1
	Greater London	5
	Hampshire	2
	Kent	2
	Leicestershire	1
	Northern Ireland	1
	Staffordshire	1
	Suffolk	1
	Surrey	1
	Yorkshire	1
UK not England	**Scotland**	2
	Wales	1
	Ireland	1
Outside UK	**Australia**	1
	Italy	1
	Total	**26**

9.1 Patient demographics of sample 2

Of 204 pairs of questionnaire booklets sent to practitioners, n = 105 were completed and returned by patients, this is a response rate of 51%. Baseline (T1) n = 105, time point 2 (T2) n = 105.

Table 33 shows the age/gender comparisons for the sub-study. The majority of this sample were women between the age of 30 years and 68 years. N=83 (79%) of sample 2 female and 22 (21%) male.

Table 33: Sample 2, age by gender

		Gender		
		Female	Male	Total
Age	30-39	1	0	1
	40-64	11	2	13
	65+	60	12	72
	Missing			19
Total		72	14	105

9.2 Number of CST sessions undertaken and year of first CST session

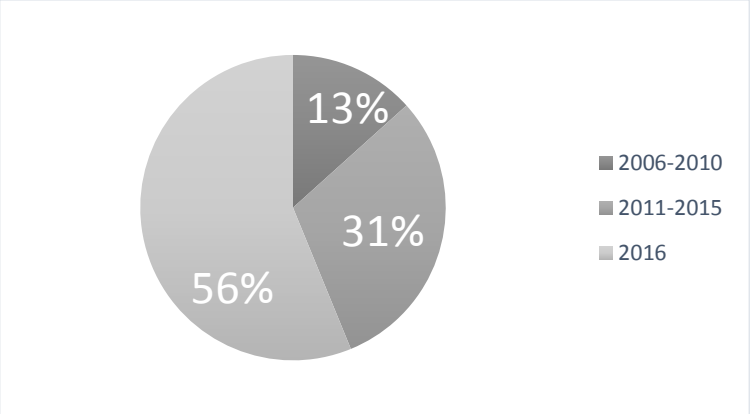
Table 34 shows that eight participants (8%) in this sample were undertaking their first CST session at baseline. Half of the sample (50%) had undertaken between 1 and 12 sessions at baseline.

Table 34: Sample 2 number of CST sessions undertaken by gender

		Gender		Total n
		Female	Male	(%)
Number of CST sessions	0	7	1	8 (8%)
	1 - 12	43	10	53 (50%)
	13 - 24	14	5	19 (18%)
	25 - 52	9	1	10 (9%)
	53+	4	3	7 (7%)
	Missing			8 (8%)
Total		77	20	105

Figure 13 shows the year in which CST users in this study had their first session. Fifty-six percent of the participants had their first session in the year this study was undertaken, 2016. Thirteen percent had their first session between ten and six years prior to the start of this study and 31% had their first session between one and five years previously.

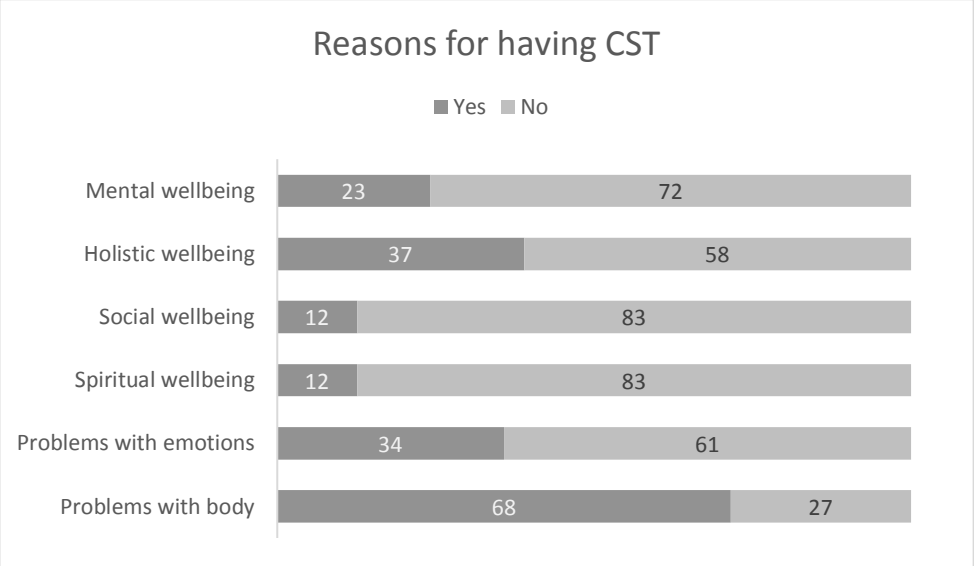
Figure 13: Sample 2, year of first CST session



9.3 Reasons for having CST sessions

Participants were asked to record their reasons for having CST sessions. They were instructed to select from the following categories which were relevant to them: problems with their body, problems with their emotions, spiritual wellbeing, social wellbeing, holistic wellbeing and mental wellbeing.

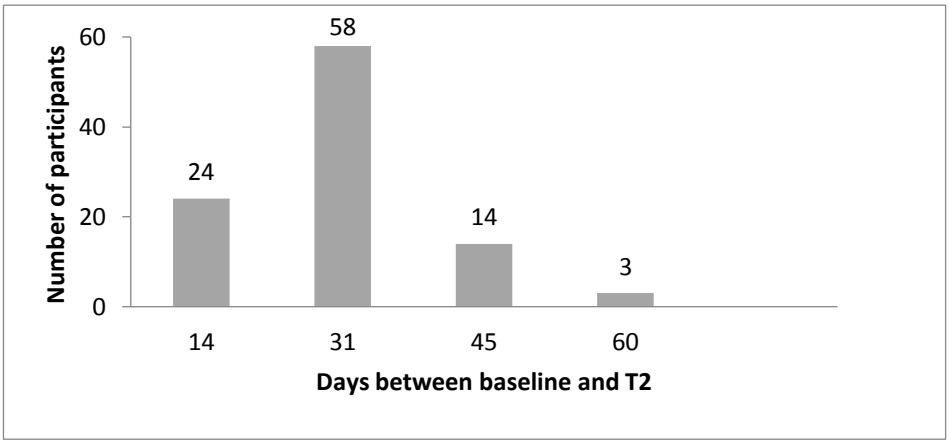
Table 35: Sample 2 (baseline) reasons for having CST



9.4 Timing of CST sessions

Figure 14 shows the number of days between CST sessions at baseline (T1) and T2. The minimum number of days between sessions is 14 and the maximum number is 60 days. The average number of days between sessions is 23.

Figure 14: Days between baseline and T2



9.5 Self-report of health status

CST users in sample 2 were asked to rate their overall wellbeing when completing the WHHQ at baseline and T: “How do you rate your overall wellbeing today?” The response options include: poor, fair, good, very good, excellent. Table 36 shows the

number of participants that rated each response option at baseline. Table 37 shows the number of participants that responded at T2.

Table 36: Sample 2, self-rated overall wellbeing by gender at baseline

		Gender		
		Female	Male	Total
Self - rated overall wellbeing	Poor	6	1	7
	Fair	29	7	36
	Good	31	10	41
	Very good	9	3	12
	Excellent	3	0	3
	Missing			6
Total		78	21	105

Table 37: Sample 2, self-rated overall wellbeing by gender at T2

		Gender		
		Female	Male	Total
Self- rated overall wellbeing T2	Poor	2	0	2
	Fair	18	4	22
	Good	35	13	48
	Very good	17	1	18
	Excellent	1	1	2
	Missing			13
Total		73	19	105

Table 38: Sample 2, changes in self-rated wellbeing between baseline and follow-up

		Self-rated overall wellbeing: Follow-up					
		Poor	Fair	Good	Very good	Excellent	Total
Self-rated overall wellbeing: Baseline	Poor	1	3	1	0	0	5
	Fair	0	13	16	3	0	32
	Good	1	4	21	8	1	35
	Very good	0	0	7	4	1	12
	Excellent	0	1	0	1	0	2
Missing							19
Total		2	21	45	16	2	105

Table 38 shows the changes in self-rated wellbeing status of the participants at baseline and at follow up. Here, it can be seen that clients fell into three groups: those who experienced a deterioration in their wellbeing between baseline and follow-up N=14 (13%); those whose wellbeing stayed the same N=39 (37.5%); and those who reported that their wellbeing had improved since baseline N=33 (31.5%); N=19 (18%) did not respond.

Over a third of participants reported that their wellbeing had stayed the same since their first session of CST and just under a third reported that they had seen an improvement in their wellbeing since their first session. It is hoped that CST had contributed to this improvement but this study was not designed to investigate that. Only a small number (N=14) reported a deterioration in their wellbeing at follow up. It is not known what events contributed to those participants who reported a deterioration in wellbeing between baseline and follow-up as this was also outside the scope of this study.

9.6 Psychometric properties of 25-item WHHQ

The measurement properties of the 25 item-WHHQ are reported in this section.

9.7 Reliability

The internal consistency reliability estimated by Cronbach's Alpha (α) for the 25-item WHHQ was 0.85 and is within the acceptable limit between 0.70 and 0.90 inclusive. Cronbach's alpha following the deletion of each item is shown in Table 38.

Table 38: Cronbach's alpha with item removed

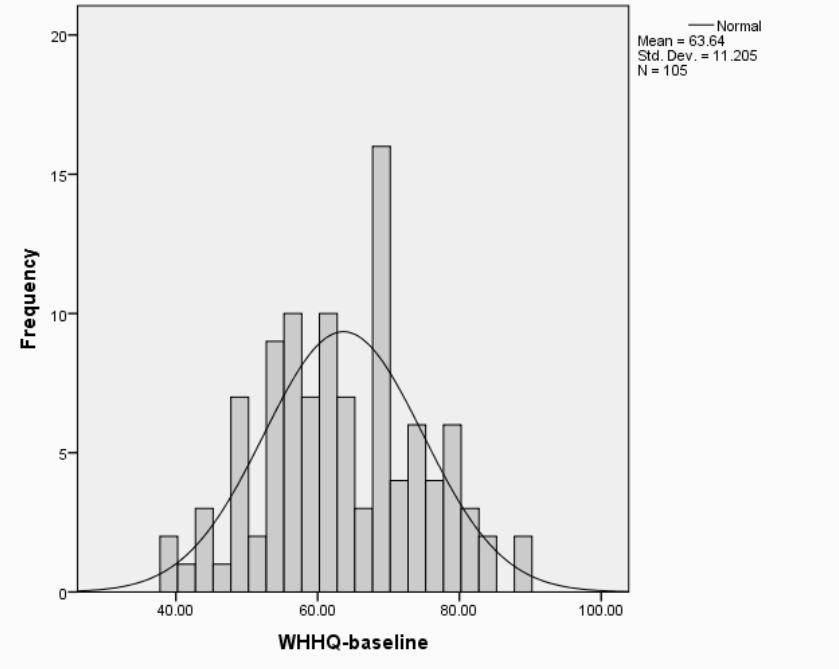
Statements	Corrected Item Total Correlation	Cronbach's Alpha if Item Deleted
Q15 My daily life has been full of things that keep me interested	.611	.840
Q53 I've felt engaged in life	.738	.836
Q47 I've felt my life has meaning	.688	.837
Q10 I've felt satisfied by my school, work or current role in life	.674	.836
Q40 I've felt joyful	.740	.834
Q39 I've felt calm	.675	.837
Q22 My awareness about my health has helped me manage life	.486	.844
Q31 I've been able to stop and reflect	.448	.845
Q24 I'm learning about myself and my body	.372	.847
Q28 I've felt in control of my health and wellbeing	.646	.839
Q2 I've felt aware of my body's needs	.535	.843
Q30 I've felt able to take care of myself	.370	.847
Q6R I've been in pain	.229	.870
Q9R My symptoms have limited my daily activities	.356	.873
Q1 I've been physically well	.451	.844
Q72R I've felt ashamed of my body	.496	.879
Q3 I've had lots of energy	.602	.839
Q62 R I've felt like a fraud	.388	.847
Q13 R I've had too many demands made on me	.173	.854
Q59 I've trusted my gut instinct	.495	.843
Q4 I've been sleeping well	.488	.843
Q42 I've felt my inner strength	.656	.837
Q54 I've felt connected to my family and friends	.620	.838
Q32 I've asked for help when I've needed it.	.387	.847
Q38 I've been able to express how I feel	.612	.839

The item numbers are for tracking purposes.

9.7.1 Distribution of the scale

The mean score was 63.6 and the standard deviation 11.2 (95% CI: 61.5 - 65.7).

Figure 15: Histogram of baseline data



In Figure 15 the data at baseline followed an approximated normal distribution, although a visual check of the scores shows a peak at around 70 points.

The scale Standard Error of Measurement was 4.32

9.8 Test of repeatability

The study design did not allow for a test of repeatability as each participant had an intervention of at least one session of CST in-between completing the baseline WHHQ and the WHHQ at time point 2. However, an anchor based method was used allowing participants to report whether their health status had changed between sessions. Table 39 shows how participants rated their health in-between sessions.

Table 39: CST users whose health remained stable in-between baseline and T2

		Frequency	Percent
	Changed a lot for the worse	2	2.1
	Stayed the same	17	18.1
	Changed a little for the better	52	55.3
	Changed a lot for the better	23	24.5
	Did not answer	11	
Total		105	

The n=17 participants whose health stayed the same have been used, to explore test-retest repeatability. A paired sample T-test was used based on n=17, to compare scores before and after and the results can be seen in Table 40.

Table 40: Paired sample T-test WHHQ baseline – T2 for those who reported no change

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	WHHQ baseline – WHHQ T2	-2.23	6.08	1.47	-5.36	.89	-1.51	16	.150

The P value associated with the t-statistic (-1.51) is 0.150 and because this is greater than the recommended p value of $p < 0.05$ the results show that there is not enough evidence to reject the null hypothesis.

Table 41: Intra-class correlation calculated model: absolute agreement

	Intraclass Correlation[b]	95% Confidence Interval	
		Lower Bound	Upper Bound
Single Measures	.801[a]	.539	.923
Average Measures	.890[c]	.701	.960

No hypothesis testing was performed but a formal *a priori* minimum level was set at > 0.7. The ICC (.801) was > 0.7, allowing the conclusion that the questionnaire has good test-retest reliability.

9.9 Responsiveness

The data are based on the 75 participants who reported improvement on the health status anchor question. The paired mean difference between baseline and follow-up T2 was estimated giving a mean change difference = 3.3 points with a baseline SD of 11.61. Of the group of participants (n = 75) who improved, the WHHQ score increased significantly, as seen in Table 42.

Table 42: Descriptive statistics 25-item WHHQ at baseline and T2 for participants who changed

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	WHHQ baseline – WHHQ T2	-3.31	8.11	.91	-5.14	-1.48	-3.60	77	.001

9.10 The effect size (ES)

For n=75 participants, the effect size was calculated in the following way:

ES = 3.31/11.61 = 0.28 (SD of change = 8.11). This indicates a small amount of change (Cohen, 1988).

Table 43: Change scores at baseline and T2 per sub-sample.

Standard Deviation (SD), Effect Size (ES), Standardised Response Mean (SRM)

	Mean	SD at baseline	ES	SRM	SD of change
25-item WHHQ Baseline N=96	63.91	11.42	**0.24**	0.33	8.34
25-item WHHQ follow-up reporting improvement N=75	67.64	11.61	**0.28**	0.41	8.11
WHHQ follow-up reporting no change N=17	64.64	9.21	**0.24**	0.37	6.08

9.11 Standardised response mean (SRM)

The SRM = 3.314/8.11 = 0.41 and indicates a small amount of change. Table 43 shows the breakdown of the sample and summary of the results.

9.12 Interpretability

9.12.1 Minimal important change (MIC)

The "changed a little for the better" group (in bold italics) was used for the MIC calculation. Table 44 shows that the HEHIQ had the largest mean change score of the four PROs used, followed by the WEMWBS and the WHHQ. The SF-12v2 MCS scale performed better than the SF-12v2 PCS scale.

Table 44: Mean change score of four PROs per the health status anchor

Health Status Anchor	*n=94*	*Mean change score (SD)*				
		WHHQ	WEMWBS	HEHIQ	SF12v2	
					MCS	PCS
Changed a lot for the better	23	5.39 (9.0)	4.84 (5.6)	6.23 (9.8)	1.87 (6.81)	3.44 (8.99)
Changed a little for the better	***52***	***2.89 (7.6)***	***3.03 (5.7)***	***4.13 (8.2)***	***.83 (4.52)***	***2.87 (7.17)***
Stayed the same	17	2.23 (6.0)	2.76 (5.4)	2.61 (8.7)	.147 (8.63)	4.24 (10.81)
Changed a little for the worse	0	0	0	0	0	0
Changed a lot for the worse	2	14.5 (2.1)	12.5 (2.1)	19.5 (2.1)	7.68 (27.16)	27.65 (19.58)

Table 45 and Table 46 provide an overview of the PRO characteristics and scores at baseline and time-point 2.

Table 45: PRO characteristics and scores at baseline and T2 N=105

	WHHQ	*WEMWBS*	*SF12v2*	
Summary scores			MCS	PCS
Minimum score	0 (worst)	14 (worst)	0 (worst)	0 (worst)
Maximum score	100 (excellent)	70 (excellent)	100(excellent)	100 (excellent)
T1, means (SD)	63.9 (11.4)	49.2 (9.4)	48.35 (10.50)	44.27 (11.04)
T2, means (SD)	66.6 (11.4)	51.9 (8.9)	49.34 (10.45)	46.68 (10.87)
Change score T1-T2, mean (SD)	2.76 (8.3)	2.73 (6.4)	.99 (7.04)	2.41 (8.92)
MIC health status anchor	2.89	3.03	.83	2.87

T1 baseline, T2 follow-up, MIC minimal important change

Table 46: PRO characteristics of scores at baseline and T2 n=105 HEHIQ scales

	HEHIQ by Scale					
	HEHIQ Scale	Outlook	Energy	Health	Relationships	Emotional Balance
Minimum score	20 (worst)	4	4	4	3	5
Maximum score	100 (excellent)	20	20	20	15	25
T1, means (SD)	70.7 (12.5)	3.55 (.74)	3.3 (.83)	3.78 (.67)	3.9 (.73)	3.5 (.79)
T2, means (SD)	74.1 (11.3)	3.74 (.66)	3.44 (.77)	3.81 (.65)	3.9 (.73)	3.7 (.66)
Change score T1-T2, mean (SD)	3.40 (9.8)	0.19 (.08)	0.14 (.06)	0.03 (.02)	0 (0)	0.2 (.13)
MIC health status anchor	4.13					

T1 baseline, T2 follow-up, MIC minimal important change

9.13 Convergent validity

This section reports (Table 47) whether there are any significant relationships between the 25-item WHHQ and the three comparator measures used: WEMWBS, HEHIQ and SF-12v2 (PCS and MCS) scales. The direction of the relationships and the strength or magnitude of the relationships are reported.

Table 47: Pearson's Correlations between 25-item WHHQ and three comparators

		Correlation with WHHQ	Significance of correlation	Number of valid responses (max N=105)
WEMWBS		.739**	<0.01	96
HEHIQ	outlook	.717**	<0.01	105
	energy	.68**	<0.01	105
	health	.54**	<0.01	105
	relationships	.73**	<0.01	105
	emotional	.73**	<0.01	105
SF12	PCS	0.15	0.112	105
	MCS	0.72	0.463	105

**Correlation is significant at the 0.01 level (2-tailed).

There is a strong positive relationship between the 25-item WHHQ and the WEMWBS = 0.74, $p < 0.01$, (n=96), and a strong positive relationship between the 25-item WHHQ and the HEHIQ outlook scale $r = 0.71$, $p < 0.01$, (n=105); HEHIQ relationship scale $r = 0.73$, $p < 0.01$, (n=105) and the HEHIQ emotional scale $r = 0.73$, $p = < 0.01$, (n=105). There is a moderate positive correlation between the 25-item WHHQ and the HEHIQ energy scale, $r = 0.68$, $p < 0.01$, (n=105). There is a weak positive relationship between the 25-item WHHQ and the HEHIQ health scale, $r = 0.54$, $p < 0.01$, (n=105).

Figure 16 shows there is no correlation between the 25-item WHHQ and the SF-12v2 physical component summary scale, r (103) = .15, p = .156.

Figure 16: Plot showing 25-item WHHQ and SF-12v2 PCS scores

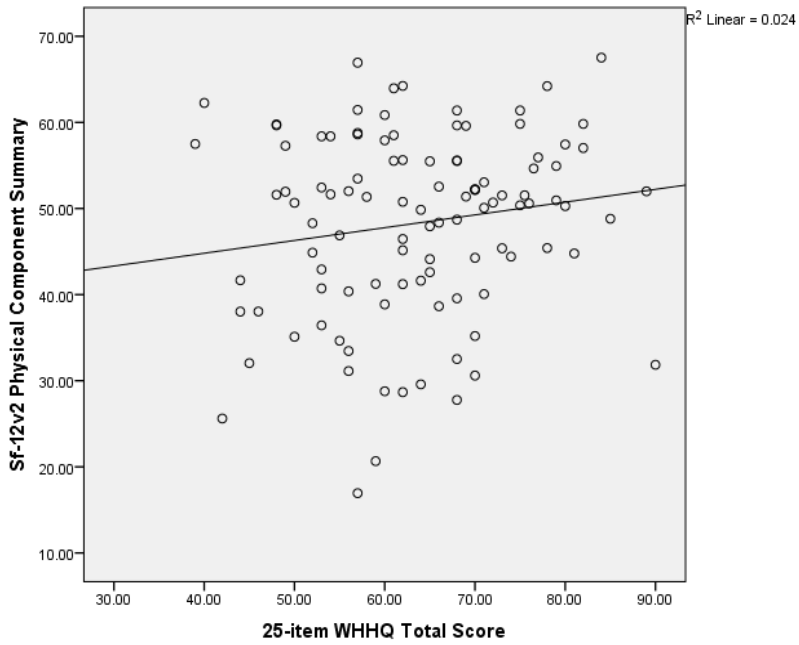


Figure 17 shows there is no significant relationship

between the 25-item WHHQ and the SF-12v2 mental component summary scale, r = (103) .072, p = .463

Figure 17: Plot showing 25-item WHHQ and SF-12v2 MCS scores

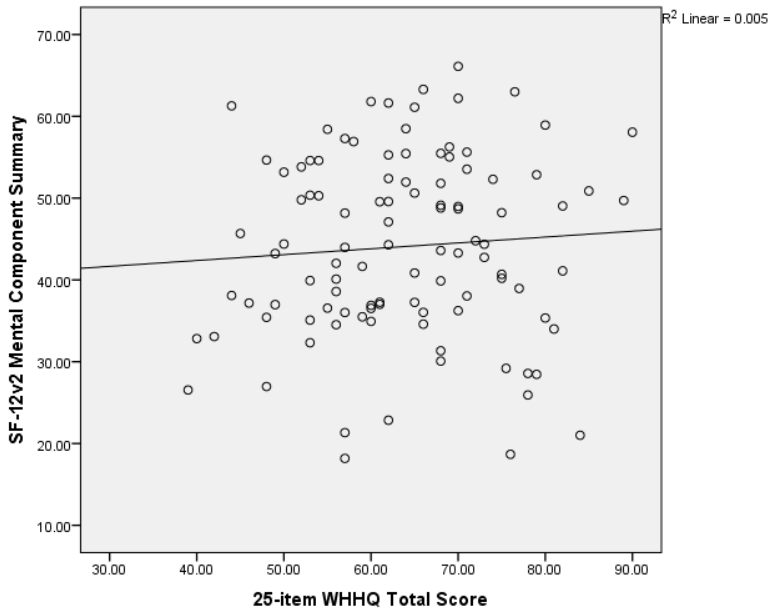


Table 48 shows the results of the effect size (ES) and SRM data for all questionnaires used in this study. All PROs show a small effect size, with WEMWBS showing the largest effect size followed by HEHIQ, WHHQ and then SF-12v2 scales.

Table 48: Effect size and SRM of PROs N=105 (SF12v2 n=97)

	WHHQ	*WEMWBS*	*HEHIQ*	*SF12-V2*	
				PCS	*MCS*
Mean change	2.76	2.73	3.40	.99	2.41
SD at baseline	11.42	9.4	12.5	10.50	11.04
ES	.24	.29	.27	.09	.21
SD of change	8.3	6.4	9.8	7.04	8.92
SRM	.33	.43	.35	.14	.27

Mean change/SD at baseline = ES, Mean change/SD of change = SRM

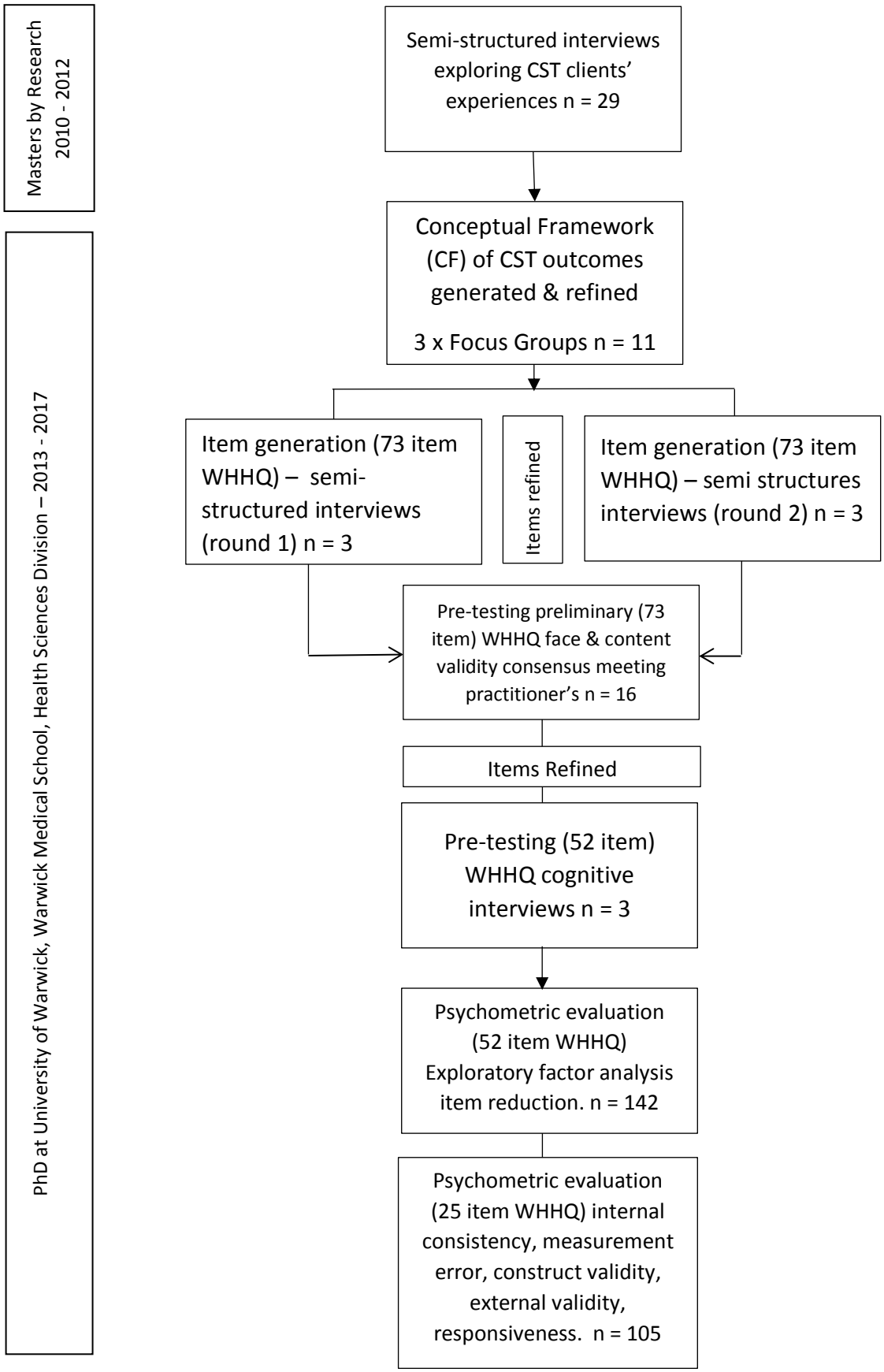
Cohen's effect size index: 0.2 is a small effect size, 0.5 is a medium effect size and 0.8 is a large effect size (Cohen, 1988).

9.14 Chapter summary

The 25-item WHHQ shows good internal consistency and good repeatability. The factor structure of the 25-item WHHQ revealed four correlated factors. The responsiveness of the 25-item WHHQ showed small effects and this is similar to the other measures used in our sample. The mean change score of the WHHQ was 3.3 and was smaller than the SEM (4.32). Convergent validity: the 25-item WHHQ performed well against the WEMWBS showing a strong positive relationship between them, and had a strong positive relationship to the HEHIQ outlook, relationship and emotional scales, a moderate positive correlation with the energy scale and a weak positive correlation with the HEHIQ health scale. A weak correlation between the SF-12v2 PCS scale and no relationship between the SF12v2 MCS sale and the WHHQ was shown. Effect sizes were comparable across all PROs except the SF-12v2 MCS scale, which was smaller.

9.15 Summary of results

Figure 18: Phases in the development & psychometric evaluation of the WHHQ



9.15.1 Conceptual framework of CST outcomes

A conceptual framework of CST outcomes was developed and qualitative work was carried out to evaluate the conceptual framework. The FDA (2009) guidance on how to develop a PRO was followed as this was deemed to be best practice. The first part of this process was to develop a conceptual framework. Having a conceptual framework lays the foundations for identifying how CST users perceive CST impacts their lives and results in a PRO to capture areas of importance.

The lack of literature related to the topic of CST outcomes hampered the development of the conceptual framework of CST outcomes. Development depended on previous qualitative work (Brough, 2012) to inform and provide the foundations for the content. The work of Brough (2012) identified that CST appears to help people with a diverse range of illnesses and others with non-specific health problems to a level of improved wellbeing. Changes in health sense were seen which combined improvement in mind, body and spirit, with participants reporting *"new levels of awareness leading to changes in self-concept; new awareness of mind-body-spirit links; and a greater awareness of their emotions. [Participants] had changes in perspectives that led to them adopting new coping strategies; they started to undertake self-care and noted improvements in their interpersonal relationships"* (pg. 124).

The layout of the conceptual framework, the domains and the relationships between them and how the components which make up each domain are operationalised within CST user's lives, was improved by using focus groups with CST users to evaluate and refine the conceptual framework. The conceptual framework of CST outcomes was endorsed by expert opinion with focus groups of CST practitioners.

9.15.2 Changes to the conceptual framework made through consensus

The feedback and input from CST users and CST practitioners was crucial to the process as, on their recommendations, important refinements were made.

These included expanding the social wellbeing domain to include 'making use of a support network'; 'feeling connected to local community' and 'intimate relationships'. 'Responsibility for self' was adopted to define self-care better in the sub-domain of mental wellbeing as proposed by CST users. The components linked to this sub-domain were separated into two 'allowing a nurturing experience' and 'taking reflective down time'. A component of 'autonomy' was also introduced. The distinctions between giving (oneself time-out) and receiving (allowing a nurturing experience) were deemed as important to CST users as these two activities indicate a choice has been made (a process of inner reflection about what one might need has taken place) and there is a call to action, for example, to book a CST session. Autonomy is required to enable an individual to make choices like these. Choosing to engage in these activities of self-care can be an important indicator of how successfully an individual may, or may not be in managing their health.

Another important refinement made based on the discussions with CST users was that arrowheads be assigned to both ends of all arrows on the conceptual framework diagram to depict the multi-directional possibilities of change. The arrowheads indicate some of the important and most common relationships between the domains, sub-domains and components as seen by CST users. How each item on the conceptual framework related to the other parts will have important implications for how a questionnaire, chosen to capture CST outcomes, performs. From the perspective of a truly holistic view of health, there is no order or hierarchy in the way in which outcomes manifest.

9.15.2.1 Topics of debate in relation to the conceptual framework content

Topics of debate arising: CST practitioners debated whether the domain of 'spiritual wellbeing' was relevant. They felt some key words were missing such as resilience and vulnerability but understood that the CST user feedback would influence such inclusions. CST users debated the relevance of 'finances' and 'sexuality' for inclusion on the conceptual framework, as these topics ignited contentious discussion. 'Finances' was excluded to ensure good face

validity and it was felt that 'intimate relationships' could include 'sexuality' and the latter was, therefore, omitted.

To recap, the conceptual framework has health and wellbeing as the central concept made up by domains (and sub-domains) of physical functioning (mobility, posture and function), mental wellbeing (responsibility for self, self-concept, emotions/feelings, mental outlook), social wellbeing (engaging in life) and spiritual wellbeing (being present); the sub-domains, energy and symptoms are presented in a different manner than the latter domains, as they can manifest in any area of an individual's life. Also, central to the conceptual framework is the theory that 'fostering self-awareness will change the way individuals relate to the different aspects of self (mind, body, spirit), others and their environment'. This is important because, as an individual becomes aware of their own process, new perspectives on how they operate in the world are seen and opportunities arise to make new choices about how they take care of themselves. The domains in the conceptual framework include some constructs included in generic HRQoL models (Fayers & Machin, 2007), together with components that extend understanding of health and wellbeing.

This is the first conceptual framework of its kind to map the range and nature of impact CST has on the health and wellbeing of CST users. CST practitioners who took part in the evaluation of the conceptual framework reported that the model has value in helping CST users and other healthcare professionals understand the scope of CST. Clients with different health states and ages were included to ensure that the conceptual framework captured the range of views of as many different CST users as possible. However, the sample was limited to English speaking participants. The conceptual framework provided a structure to assess the content of any potential measures identified during the literature searches. It also provided the basis for the development of a new PRO to assess change in the health and wellbeing of those having CST. Two components of the conceptual framework are new to CST evaluation: 'developing self-awareness' and 'taking responsibility for self' (Brough, 2012). Both components are fundamental to health, in its broadest sense, as good

health requires patient participation and CST users report that having had CST they develop an ability to take responsibility for their own health (Brough *et al*., 2015).

Creating a conceptual framework is an art not a science and the conceptual framework is an evolving document which is still under discussion and may change further as others become interested in these topics. On reflection, the components linked to symptoms e.g.: recovered, relief sustained/temporary, and severity and frequency in the future may be removed all together as these components only represent a small part of the overall conceptual framework of CST outcomes. Alternatively, they could be expanded to incorporate the dimensions of intensity, frequency, duration, nature, impact and bother. Evaluating symptoms in this way is similar to disease models used in generic PRO evaluation.

9.16 The literature review of PROs developed for CST and CAM

Three systematic searches were undertaken, the first was done to identify whether any PROs had been developed specifically for CST and this search yielded no results, suggesting that a therapy-specific PRO for CST does not exist.

A second systematic search was carried out to identify and appraise PRO measures which have been used in CST evaluation studies. Eight studies were identified and twenty-five outcome measures had been used to assess primary and secondary outcomes within these studies. Of these, the majority were disease specific measures and, whilst they were suitable in the context of the studies that used them, they do not fulfil the conceptual framework criteria. The others assessed pain intensity, pain acceptance, sleep quality, perceived stress, body awareness, subjective physical wellbeing, impression of improvement, kinesiophobia and disability. Some CST studies had used PROs from other sectors of healthcare and these were examined. Applying the criteria of holism, HRQoL questionnaires are more applicable for CST as they cover multiple domains. Three HRQoL questionnaires were identified and

appraised to see if they capture the outcomes of CST users, the SF-36, SF-12 and the EQ-5D. The suite of measures to which the SF-36 and SF-12 belong, are well validated across a diverse range of healthcare settings and are highly regarded and widely used to capture generic patient reported outcomes. Yet, both lack the ability to detect how a person may be adapting to their situation or their ability to cope. They do not have items to evaluate self-awareness or self-care and do not have items to capture spiritual outcomes. The EQ-5D does not capture outcomes of a spiritual or social nature. With only one item relating to mental health and three items on the physical domain it is unbalanced across the two domains. Its main advantage is that is can be used to calculate QALYs and, thus, address cost utility. None of the three measures met all the criteria required for CST.

Only four of the eight studies reviewed assessed outcomes beyond a single disease/condition 1) CST has mostly been evaluated to assess the extent it impacts on medically defined conditions and 2) There has been less interest in the full impact CST has on health and wellbeing (Brough, 2012). Current quantitative studies of CST, thus, do not do justice to CST practice.

The aim of the final search was to identify and appraise any PROs developed for CAM more generally. Since CST shares some characteristics like holism and attention to spirituality, it was possible that measures for other CAM might work for CST. Four questionnaires developed specifically for CAM populations were identified: Complementary and Integrative Medicine Outcomes Scales (CIMOS) (Eton *et al.*, 2005), the Measure Yourself Medical Outcome Profile (MYMOP) (Paterson, 1996b), Harry Edwards Healing Impact Questionnaire (HEHIQ) (Bishop *et al.*, 2010), and the Self-Assessment of Change (SAC) (Ritenbaugh *et al.*, 2011). CIMOS matched the Conceptual Framework of CST outcomes best from a conceptual perspective but has a high emphasis on symptoms such as pain and fatigue which are more akin to the biomedical view of health not CST. Also, reliability and validity are limited to the studies reported during development, with limited methods being applied with small sample sizes. The CIMOS domains were developed for use in a modular way,

making comparisons across studies difficult and sensitivity to change is yet to be established.

The MYMOP lacks provision for items of a spiritual nature and, as they were important to CST users, the ability to capture such outcomes is essential. MYMOP is unable to measure change in new problems and similarly, in several co-existing problems and concerns, failing to capture unexpected change in new symptoms or symptoms beyond the two identified by the responder as important. The MYMOP has moderate reliability, good validity and is sensitive to change (Paterson, 1996a), and, whilst its individualised nature is patient centred, using this type of measure brings its own challenges for example, measuring the same condition with a homogenous population would be difficult and comparisons across studies unworkable.

HEHIQ was developed in a population like that of CST. The contents are a good match despite the items related to physical domains being limited and there being no items capturing self-care, self-awareness or patient engagement. The HEHIQ's reliability and validity is only evidenced by the development papers and, whilst the HEHIQ was shown to be sensitive to change on some domains during validation (Bishop *et al.*, 2010), further work is required to establish its responsiveness.

The SAC lacks items on self-care, self-awareness and patient engagement. Whilst conceptually the SAC is a good match, reliability and validity are hampered by the fact that the pooled data from seven studies in different settings used to undertake the psychometric analysis had different versions of the SAC prepared depending on the setting (some used paper and some used online). This may have influenced the way people responded and may have presented challenges and data entry burden due to the paper versions being VAS responses. In addition, sensitivity to change has yet to be established see Table 12, pg. 73.

An additional measure WEMWBS (Tennant *et al.*, 2007) was appraised, although WEMWBS is not specifically developed for CAM, a strong case has

been made that it is likely to be useful in this context (Stewart-Brown, 2015a) and it covers a domain of importance for CST which was not covered by other measures which is mental wellbeing. WEMWBS is reliable and very widely used; it has good validity and has been shown to be responsive at both group and individual level (Maheswaran *et al.*, 2012), see Table 11, pg.72.

9.16.1 Results: qualitative

In parallel to the development of the conceptual framework, a list of 73 items was generated for the draft PRO. Some items were generated using verbatim statements from previous qualitative work carried out by Brough (2013) when a large number of people were interviewed and the outcomes they attributed to CST were identified. Other items were created based on outcomes of importance identified in literature cited by Brough (2012). Items were mapped to the draft conceptual framework to ensure that there was at least one item corresponding to the components on the conceptual framework. Much consideration had been given to the layout and instructions of the questionnaire based on observations of other questionnaires and taking the literature on PRO development into account (De Vet *et al.*, 2011; Fayers & Machin, 2007; Streiner *et al.*, 2015). A draft PRO was developed and two rounds of semi-structured interviews with CST users and a consensus meeting with CST practitioners were carried out to determine the item selection, layout, instructions, recall period, response options and questionnaire name.

During this process items were removed if they were found to be unclear, difficult to understand or confusing, or duplicated. Items were revised to improve understanding or to ensure tense consistency. 21 items were deleted during the evaluation, 36 items (1 item was revised after round 1; 26 items were revised after round 2; and 9 items were revised after the consensus meeting with CST practitioners) were revised to improve the comprehensibility of the statements. There were 52 items remaining for inclusion on the WHHQ.

The 52-item WHHQ was pre-tested using cognitive interviews with a small group (n=3) of CST users to assess face and content validity. No changes were made to the content after round 3, but the response options were revised to 'all or most of the time' based on the client reports.

9.16.2 Results: quantitative

The measurement properties of the 52-item WHHQ were assessed in a sample of CST users. A summary of the sample demographics is described here. N = 142 CST users took part in this study, they were a heterogenous sample and were located throughout the UK. Of the participants in this study 81 % of them were female and 19% were male, they had an age range between 16 years and 83 years. Participants had received different numbers of sessions over different periods of time; but over half (52%) of the sample started sessions during the year in which this study was carried out, 2015. Participants were asked to self-rate their overall wellbeing at the time of completing the questionnaire: 47% rated as good and 3% as poor.

Construct validity was considered in relation to the conceptual framework of CST outcomes and as a means of organising the items and presenting the 52-item WHHQ draft questionnaire to participants; not in terms of actual domains or factor structure, as all domains and the items on the conceptual framework are intrinsically linked. Exploratory factor analysis was undertaken as an item reduction technique. During this process, 32 items were deleted due to multiple factor loadings or weak communalities. As a result, 19 items remained on four factors. The qualitative data were cross checked to ensure good face and content validity, and eleven key items of importance to CST users were not been retained during the EFA process. A consensus meeting with CST practitioners re-evaluated these items for inclusion on the WHHQ, six of which were returned to the item pool for further evaluation, five were considered not 'sensitive to change' or found to be 'too subjective'. No consensus was reached with item 75, as there was concern it could influence response rates. The modifications were made and the result was a 25-item WHHQ. As it was hypothesised that the factors correlated, an oblique rotation was applied to the data. The measurement properties of the 25-item WHHQ are summarised next.

9.16.3 Evaluating the measurement properties of the 25-item WHHQ

A psychometric assessment of the measurement properties of the 25-item WHHQ was undertaken in a second sample of CST users. CST users completed a 25-item WHHQ before a session of CST (baseline) and before a follow-up session of CST (timepoint 2). N = 105, 79% of sample 2 were female and 21% were male. The participants had an age range between 30 years and 83 years.

The following measurement properties were evaluated: acceptability (floor/ceiling effects and skew of scale scores), reliability (internal consistency and item-total correlations) and validity (convergent validity) and the final version of the 25-item WHHQ's ability to detect change (responsiveness) was evaluated.

9.16.4 Acceptability

The data at baseline appeared to be following a normal distribution, although a visual check of the scores shows a peak at around 70 points on the WHHQ. This artefact in the data could be due to participants 'yay-saying' or answering 'often' down the page. No floor or ceiling effects were shown.

9.16.5 Reliability

25-item WHHQ had an α of 0.85, the mean is 63.6 and the standard deviation is 11.2. The alpha value suggests good internal consistency for the 25-item WHHQ as the alpha value exceeds the suggested minimum of 0.07. The true population mean on the WHHQ for this sample of CST users lies between 61.5 and 65.7 (CI 95%).

Repeatability was not formally evaluated because participants were having CST sessions in between baseline and T2. However, a sub-group who reported that their health status had 'stayed the same' between baseline and T2 was reported on (n = 17) and an intra-class correlation (ICC) was calculated as 0.801. This exceeds the recommended value of 0.7 and, therefore, we can conclude that the questionnaire has good test-retest reliability.

9.16.6 Validity

To establish convergent validity, three comparator measures were used in this study. Measures were selected based on their content and face validity, reliability and validity. Three comparator questionnaires together cover most of the outcomes important to CST users. The HEHIQ, whilst its reliability and validity were limited, was a good match conceptually and had been validated in a population similar to that of CST. WEMWBS captures mental wellbeing as an important domain for CST users and has good reliability and validity. The SF-12 has provision for mental and physical domains and is well validated. It was hypothesised that a strong correlation ≥ 0.70 in a positive direction would be seen with all comparator measures and their domains. It was found that there was a strong positive relationship between the 25-item WHHQ and the WEMWBS; the HEHIQ outlook; relationship and emotional scales; a moderate positive correlation with the HEHIQ energy scale and a weak positive correlation with the HEHIQ health scale. A weak correlation between the SF-12v2 PCS scale and no relationship between the SF12v2 MCS sale and the WHHQ were shown. This may be because both the SF-12v2 PCS and SF12v2 MCS items are symptom focused unlike the 25-item WHHQ. This point will be expanded on in the discussion section.

9.16.7 Responsiveness

Responsiveness was evaluated using a global rating scale as an external anchor, with n = 75 reporting an improvement. The standard error of measurement (SEM) is 4.32. The paired mean difference between (N=75) baseline and follow-up T2 is 3.3 points with a baseline SD of 11.61.

The effect size shown by the 25-item WHHQ = 0.28 and shows a small amount of change based on Cohen's effect size index (1988), and the standardized response mean = 0.41. Although small, effect sizes were comparable across all PROs except the SF-12v2 MCS scale. The results are discussed in the next section.

10 Discussion

10.1 Chapter Overview

This section presents a summary of the findings from this study and discusses their implications. Adequacy of the research methods applied and the implications for instrument validity are discussed. The section also describes how these results link to the reviewed literature described earlier in this thesis (pg. 44) and addresses the content of the conceptual framework of CST outcomes including aspects such as HRQoL, wellbeing, taking responsibility for oneself and the social aspects of health. The contribution this research offers to the field of CST and PRO development is presented in addition to the strengths and limitations of the study. The section concludes with options for future research.

10.2 Summary of the findings

This thesis set out to develop a PRO that was suitable for use to evaluate CST primarily in a clinical setting but also for research. Therefore, simplicity of layout, short completion time and easy scoring were prerequisites. The aim was to create a unidimensional scale covering the multiple domains of health and wellbeing impacted by CST, described in a conceptual framework and providing a single score calculated by adding the item scores.

The starting point for this research was a qualitative study reporting clients' experiences of CST (Brough, 2012) undertaken before this thesis started. The data collected in the latter study and the analysis, provided the basis for development of both the conceptual framework and the questionnaire items in the research reported here. The conceptual framework is important because it is the first of its kind for CST and it maps the domains of health and wellbeing impacted by CST which are not addressed by existing PROs such as the 'development of self-awareness' and 'taking responsibility for self'. As some of the latter are also impacted by other CAM the PRO developed during this thesis may have wider applications.

This thesis began with a literature review to ascertain whether the proposed research had been attempted before. The review was repeated at the end of the study to include research published since embarking on the thesis. Candidate measures were identified and assessed for suitability against the conceptual framework and their psychometric properties reviewed. A gap in the literature was identified as none of the measures covered all the components on the conceptual framework but measures suitable for assessing construct validity were identified.

The draft PRO was developed and evaluated iteratively in a series of qualitative and quantitative studies. The final version of the WHHQ comprised 25 items, 80% of which are positively worded, covering holistic health (physical wellbeing, mental wellbeing, spiritual wellbeing and social wellbeing). Validity and reliability were better than, or comparable to, that of other measures used in our study. It was well liked by practitioners and clients of CST who felt the measure had good face validity. It showed anticipated convergent validity with WEMWBS and HEHIQ; weaker correlations with SF12v2 PCS scale and none with SF-12v2 MCS scale. It met the psychometric criterion of reliability, with good internal consistency. Repeatability over time was good and it was responsive to change taking place during treatment. However, as the mean change over the course of treatment was smaller than the SEM assessed in the repeatability study, it may be that the WHHQ cannot distinguish between the error of measurement and health improvement. Further research is needed on this with larger, more diverse populations. In this study sample, some participants were long term users of CST in which maintenance of good health was considered a positive outcome. Only two participants reported deterioration making it impossible to test responsiveness to deterioration adequately.

10.3 Adequacy of research methods and implications for instrument validity

10.3.1 Literature review

The literature searches failed to identify measures specifically developed for CST, and further searching for measures developed for CAM in general failed to identify

any that covered all the aspects of the conceptual framework. This may be because measures which cover the conceptual framework have not yet been developed, but it could also be due to methodological issues. For example, studies on measurement properties are sometimes poorly reported and consistent terms are not always used in titles or abstracts. This area of work uses broad terminology for outcome measures and this makes it difficult to find relevant measures. Indexing by the National Library of Medicine can be incomplete and unpredictable. MeSH terms, whilst producing broad searches, may not be tagged with the appropriate specific terms related to this work.

A systematic review undertaken by Hunter (2013) to identify patient questionnaires for use in integrative medicine was helpful in refining the searches, suggesting possible search terms and identifying sources of potential measures. The 2017 searches for this thesis identified a paper reporting a PubMed search strategy for studies of measurement properties of PROs (Terwee et al., 2009). This might have been useful in the preliminary searches but did not come to light until the later stages. The 2017 searches also identified an additional CAM PRO known as the SAC. This might have been considered as a comparator measure to assess construct validity if it had come to light at an earlier stage.

10.3.2 Development and validation of the conceptual framework

Sequential mixed methodology has lent itself to the iterative nature of PRO development. The qualitative methodology allowed for exploratory enquiry in the development and evaluation of the conceptual framework, item selection, PRO design and layout and quantitative methodology enabled empirical evidence to be gathered to support factor structure and to establish psychometric properties. The conceptual framework and the initial list of 35 items were developed simultaneously. The 35 listed items (which were verbatim statements from CST users based on research findings (Brough, 2012) and relevant measures) were examined to see the extent to which they mapped onto these items. Looking to see what others have done can save work and be a way of identifying items that have been previously tested and are psychometrically sound (Streiner & Norman, 2008).

The evaluation of the conceptual framework in focus groups ensured the viewpoints of both CST users and CST practitioners had been taken into account; and enabled the conceptual framework to become the basis of a theory covering all possible effects of having CST. Whilst it would have been ideal to have a more equal gender split amongst the CST user focus groups, only female CST users took part. The gender distribution of this sample was expected to be mainly female, as this is the distribution of consumers of CAM treatments in the UK (Hunt *et al.*, 2010.)

Intergroup heterogeneity was realised by having both CST users and CST practitioners' focus groups done separately. Group discussion provided much insight and improvements to the overall layout and content of the conceptual framework. The conceptual framework was used to ensure that items in the original pool (35 items) covered all domains of health and wellbeing deemed important to CST users. Streiner and Norman (2008) suggest that each theme be represented by at least one item. Items were developed for the WHHQ from items in existing PROs which evaluated the missing domains and added to the original list bringing the total to 73. Thus, more than half of the items on this list were constructed through modifying items from other sources to suit the needs of this population. However, no items on the draft list (73-items) were taken verbatim from other sources. In reflection, this was a missed opportunity to include items that had been validated in other populations and which may have been suitable for this population too. For example: identifying a sleep item from another questionnaire may have enhanced the validity of this item on the WHHQ. CST user and practitioner input during development of the conceptual framework was invaluable in enhancing the content in this way, extending the selection of items beyond the 35 verbatim items from Brough's study (2012).

10.3.3 Development and validation of the WHHQ

The next steps followed the sequential mixed methods study design, which allowed the advantages and integrity of each method to be upheld as each sub-study was completed before the next one began. Tension between the qualitative and

quantitative approaches did arise during the research process at the point when the first part of the quantitative data had been collected and subjected to EFA as an item reduction technique. Revisiting focus group data at this stage suggested a potential but conflicting choice of whether the criteria for item deletion be maintained by presenting a questionnaire with fewer items. Alternatively, should the face and content validity be upheld while honouring the suggestions that CST users and CST practitioners had made in the early stages of questionnaire development? The decision to reinstate items of importance to CST users was made to maintain face and content validity. These items were then subjected to further quantitative testing, some were subsequently retained and some deleted. Having obtained the engagement of CST users and CST practitioners as experts as part of the development process, meant that face validity was good and the decision to honour their input when possible, enhanced the potential acceptability of the questionnaire for clients and practitioners. An opportunity was missed at this stage to reconsider the wording of the items which the quantitative analysis suggested should be deleted. For example: item 4, 'I've been sleeping well' was the first item to be deleted during the quantitative analysis yet, both CST users and CST practitioners felt that 'sleep' was a good indicator of health status. It was found that users quite often referred to 'having slept really well' since their last session, or that they had not been sleeping well and they linked this to circumstances or health issues when recalling how they have been since their last session of CST. Changing the wording of these items may have produced an item that was both psychometrically sound and which contributed to face validity.

The items of the 25-item WHHQ do not fully represent the components of the conceptual framework, as some items did not enter the final version of the WHHQ due to the item reduction process. The components not represented include 'living my core values', 'connection to self, divine, animals and nature', 'trust, faith in life, having compassion', 'feeling connected to local community' and 'adopting new attitudes or coping strategies'. As these components represent core values which are fundamental to wellbeing, further work is required to refine and test the items

mapped to these areas of the conceptual framework to fulfil all outcomes of importance to CST users.

10.3.4 Determining the format of items and response options

Much deliberation was given to the format of the items and response options and the final layout of the WHHQ. Participants in the focus groups liked the simple presentation. In earlier versions of the draft (52-item) WHHQ when there were large numbers of items, participants liked the sub-headings indicating the different domains covered, reporting that it helped them navigate and think through the process of completing the questionnaire. As the total number of items was reduced it became possible to get all items on one side of A4 paper but not the sub-headings. Having all items on one side of paper reduces incomplete responses due to responders not turning over the page and, thus, fulfilled one of the goals of the WHHQ development.

In regard to response options, a direct estimation method was applied "*designed to elicit a direct quantitative estimate of the magnitude of an attribute*" (Fayers & Machin, 2007) in the form of a Likert scale (Likert, 1952). Likert scales are bipolar and the descriptors commonly describe agreement with a statement (e.g. disagree strongly – agree strongly). On earlier versions of the WHHQ, different terms for the middle point (sometimes, don't know, does not apply) were tested to ensure that the middle amount of the attribute was reflected and not the responder's inability to answer the question. It was recommended by participants that the response option 'all of the time' be changed to 'most or all of the time' to avoid end-aversion bias. End-aversion or central tendency bias refers to the reluctance of some responders to rate at the extremes of a scale. On reflection, revising the absolute response statement of "never" to 'almost never' may have been a more balanced approach, despite this not being identified as an issue by the participants.

Having input from CST users and CST practitioners helped to inform the layout and format of items and the selection of response options considered. This ensured that participants gave the appropriate answers to enhance optimising (the ideal way of

responding to an item) and limit satisficing (giving an answer which is satisfactory, but not optimal) (Krosnick, 1991). The participants endorsed the final layout, format of items and response options.

10.3.5 Efforts to limit response bias

The simple design and easy administration of the new PRO is a benefit but bias in responses can occur. To diminish 'yay-saying' bias - the tendency to give positive responses (Couch & Keniston, 1960), five of the items on the new PRO were reverse scored to ensure that the responders did not give the same answer down the page in a straight line. This gave rise to 80% positive items *versus* 20% negative items on the 25-item WHHQ. This can, however, place additional cognitive demand on the responder (Fayers & Machin, 2007) and, during the cognitive interviews participants reported that the shift in direction of the item responses required a little re-orientation on their part when marking a response. With hindsight, all of the items could have been positive to negate this burden. The best approach is easily testable with further research by presenting all items as positive responses against the current version 25-item WHHQ (80% positive and 20% negative).

10.3.6 Measurement theory applied in this study

In this study CTT was applied and proved the best tool to use, as the requirements were met for field testing the WHHQ. The CTT methods required simple mathematical analyses and the model parameters estimation was conceptually straight-forward, aiding the process. Furthermore, CTT has weak assumptions which were easy to meet with the theoretical model. By contrast, the underlying assumptions required for IRT were not met. This was partly because the sample size was smaller than anticipated - IRT methods require much larger sample sizes (typically over 500) (Hambleton & Jones, 1993). However, more importantly, when using IRT, items are defined so they have a hierarchical order of difficulty and this hierarchical assumption does not apply to the domains of impact of CST. For example, there is no suggestion in the conceptual framework that mental wellbeing is more or less important than physical wellbeing.

10.3.7 Determining the scoring

The 25-item WHHQ is scored by summating all the item totals (0-4) together to create one total score (0 – 100) to make the scoring as simple as possible. Despite this, it was noted that, during the data entry for the final tranche of quantitative data, some practitioners had still made errors in adding item scores. Discussions with CST practitioners clarified that, whilst some lacked confidence in calculating the scores, others did not have the time during a session of CST to complete the scoring. Given that some practitioners found it difficult to score even the simplest instruments at the same time as doing clinical work, if the instrument is going to be valuable in the clinical setting as intended it would be useful to investigate and trial electronic aids which could calculate scores automatically.

10.3.8 Semi-structured interviews

The geographical spread of participants was so great that it was not possible, due to participant travel and financial constraints of the study, to gather all participants together in one place to convene a focus group. Therefore, the decision to undertake semi-structured interviews instead of focus groups was made. Whilst this was more time consuming, the researcher was given the opportunity to glean greater insight into why items on the 73-item WHHQ were selected or excluded.

10.3.9 Data collection (qualitative)

Having had previous qualitative research experience, the researcher's listening and attention skills aided the focus groups and interviews, allowed her to meet the research objectives and allowed participants to share what was important to them.

10.3.10Organising and analysing data

The chosen study design and mixed methods approach gave flexibility to select methods best suited to organising and analysing the data.

Qualitative

The focus group data were dense, so organising the data for each round to enable comparisons to be made across the groups was a challenge. The use of inductive thematic analysis to identify, analyse and report themes within the data worked well to enable this process. Documenting the revisions made to the conceptual framework and the draft PRO were key to ensuring coherency when condensing the data for reporting.

Quantitative

SPSS was the software chosen to organise and analyse the data as it met the needs of this study. Setting up and managing the databases in the correct manner and ensuring variables were organised correctly for both samples were both instrumental to the smooth running of the analysis during each phase of the PRO development ensuring continuity and flow. There was little missing data in this study, none of the cases had more than 5% of the items missing. This suggests that the WHHQ's face and content validity are good.

10.3.11 Reflexivity

In this section I reflect upon the strategies I used to enable reflexivity and share some of the personal challenges I faced during this study. I undertook monthly academic supervision, which became an important resource for me. Academic supervision was a useful place for me to discuss the practical aspects of the research process and to keep me focussed. It was also an opportunity for a reality check and some feedback from experienced researchers who had trodden the path before me. On many occasions my lack of confidence in my own ability to learn, apply and understand statistical and psychometric methods that I felt were beyond me, would engender challenging emotions, 'imposter syndrome' and the fear of failure, despite the best efforts of my supervisors to assure me that I had the knowledge and had applied the methods correctly.

I had opportunities to present my research at conferences throughout the study (Brough *et al.*, 2016; Brough, 2014) and, if I was presenting in to an academic

audience, I would go into a blind panic and have enormous anxiety about the experience. At times this was debilitating, giving rise to a freeze response within me which prevented me from getting on with the research, sometimes for weeks at a time. My academic supervisors supported me through this process by ensuring that I knew my material and had opportunities to practice delivering the presentations and to prepare myself. On reflection, what became clear to me towards the end of this study was that I had not identified or owned the aspect of myself that was the 'academic/researcher', and feeling like an' outsider' in the university environment became the norm. Working from home and not having regular connection to fellow PhD students or being based within the university culture amplified the feelings of isolation and not fitting in.

By contrast, when disseminating my research to CST practitioners in person or via articles within the CSTA magazine, I felt enthusiastic, confident, clear about the material and the message I was conveying at the time. I used lay language mostly, but would use the appropriate terminology selectively to educate the CST community over time. I welcomed questions and comments with no concerns about whether I might be able to answer. I always felt energised, and my sense of purpose was reinforced through these exchanges. Within these settings I felt very much like 'one of them' part of the CST community and an 'insider'.

Keeping a field work journal enables researchers to keep track of assumptions and emotional reactions (Gilbert, 2000). Whilst this is something I did during times of stress, for me, a more useful process was engaging in personal psychotherapy frequently as a self-care strategy. It was in these sessions that I could talk freely about the discomfort, pressure, tensions and vulnerabilities I felt on a personal level about being a PhD student and undertaking a research project which had multiple stakeholders and required attention and energy to the exclusion of my other responsibilities including my family. It also enabled me to explore the many occasions where I felt I had to just sit with my own anxiety about not completing the process or failing in some way until such a time when a glimmer of faith would provide the impulse to continue anyway, even if I was unclear about the methods I

was applying. I had to trust that I would be clear enough to write about them once I had completed the analysis.

For me, downtime or time not engaging with the research was important to enable me to process, reflect, rest and plan. Times away were beneficial even when the time pressure felt enormous.

What I have realised, during my reflections, is that, throughout this process, the most rewarding and energising part of the journey was meeting the many practitioners who encouraged me at different times. Although they may not have realised it at the time, their words of gratitude for the work I was undertaking or their comments about how the WHHQ would be of benefit to the field of CST sustained me on the many occasions when I was feeling exhausted and overwhelmed with the process.

10.4 Findings in relation to the literature review

This section links to some of the literature presented in the introduction of this thesis, explaining the relevance it has for the conceptual framework of CST outcomes and the new PRO.

The literature review identified two PRO typologies that cover many of the domains identified as of importance to CST. These were health-related quality of life (HRQoL) and wellbeing. The relationship of these to the 25-item WHHQ warrants further discussion.

10.4.1 Health related quality of life

The term 'quality of life' is ill-defined (Fayers & Machin, 2007). To differentiate between QoL in general and how it is applied in wider context and what is necessary for medicine and clinical trials (such as evaluating aspects of health influenced by disease or treatment), HRQoL has been used for clarity. The concept of HRQoL includes domains of physical, psychological/emotional, social dimensions of health and functioning. Some HRQoL questionnaires allow for spiritual wellbeing (Fayers & Machin, 2007) and, at first, glance HRQoL appears to be a reasonably good match for evaluation of CST. However, existing HRQoL questionnaires tend to

be linked to functional capacity suggesting that if an individual is unable to achieve full psychological, physical or social functioning that their HRQoL is somewhat diminished. The issue here is that different forms of functioning, especially physical functioning, can be regarded as determinants (Fayers & Hand, 1997) of HRQoL but do not necessarily reflect the true level of their HRQoL (Fayers & Machin, 2007); for example, a person in a wheelchair can achieve a high level of wellbeing. Measuring functional status assesses whether there are problems that could be detrimental to HRQoL but does not indicate how a person perceives their HRQoL to be. Other indicators of HRQoL, of relevance to the WHHQ, are personal wellbeing and life satisfaction, both of which have recently attracted much empirical research (Huppert & Ruggeri, Forthcoming 2017) and these are discussed below.

The dimensions involved in HRQoL are numerous, with instruments measuring this construct typically containing many items and often multi-item sub-scales (Fayers & Machin, 2007), making them burdensome to patients. Practitioners may find it difficult to score multidimensional questionnaires and interpret their findings. Short HRQoL measures like EQ-5D and the SF-12v2 do not cover the key domains of CST outcomes such as spiritual wellbeing or have items that tap 'taking responsibility for self' or 'increased self-awareness'.

10.4.2 Wellbeing

Brough *et al*., (2015) reported that one of their key findings was that the process of CST helped users to see the interrelationship between the dimensions of wellbeing and to relate to a more holistic health paradigm. Therefore, in the case of CST, wellbeing is important and needs to be measured. Objective circumstances influence the wellbeing of individuals but wellbeing is a subjective state and is considered in the context of how well a person perceives their life to be going (Huppert & Ruggeri, Forthcoming 2017). There is currently no consensus on how to define wellbeing. Some scholars regard happiness a sufficient definition (Kahneman *et al*., 1999; Layard, 2005; Fredrickson, 2009 cited in (Huppert & Ruggeri, Forthcoming 2017) while others claim wellbeing encompasses psychological function, not all functional ability which includes physical, sense of

meaning and positive relationships (Deci & Ryan, 1998; (Huppert & Ruggeri, Forthcoming 2017; Ryff, 1989). Others believe that the experience of wellbeing combines both 'feeling good and functioning well' (Huppert & So, 2013; Keyes *et al.*, 2002; Seligman, 2011).

To measure wellbeing, questions need to be asked about positive experiences (Huppert & Ruggeri, Forthcoming 2017). This breaks away from the traditional and disease-based models of health where the focus lies predominantly in measuring disease and symptoms. Just because an individual does not endorse symptoms on a scale, we cannot conclude that they have a high level of wellbeing. All that is shown is that they do not have symptoms of the condition being measured.

Like HRQoL, wellbeing has multiple dimensions some of which link to the outcomes of importance to CST users. Firstly, mental wellbeing will be addressed then wellbeing more generally. Examples of mental wellbeing measures that have some relevance in this context include, the WEMWBS (Tennant *et al.*, 2006) chosen as a measure of criterion validity in this thesis, Seligman's PERMA (2011) (positive emotion, engagement, relationships, meaning and accomplishments) all of which are represented by items on the 25-item WHHQ except for accomplishments. Ryff's Scales of Psychological Wellbeing (1989) the domains of which (autonomy, environmental mastery, personal growth, positive relationships, purpose in life and self-acceptance) are also represented by items on the 25-item WHHQ. Huppert and So (2013) developed 10 dimensions that define the positive end of the mental health spectrum: sense of competence, emotional stability, engagement, sense of meaning, optimism, positive emotions, positive relationships, resilience, self-esteem and vitality. There are areas of overlap between HRQoL, wellbeing measures and the 25-item WHHQ, as the latter has items which represent, in part all, of the dimensions of wellbeing mentioned in the literature. There are, however, two areas of the WHHQ which are not represented either in wellbeing and HRQoL measures. These are discussed next.

10.4.3 Taking responsibility for oneself

An important outcome reported by CST users was an increased ability to look after themselves (Brough et al., 2015). In this thesis, CST users who evaluated the conceptual framework viewed self-care as 'taking responsibility for oneself' and made clear distinctions between allowing a nurturing experience and taking reflective downtime (such as meditating) as these behaviours were regarded as important to self-care. Participants indicated that through enhanced awareness of their health they could make decisions about their needs, undertake activities to suit their health priorities and improve their overall quality of life. Similar findings have been reported elsewhere in relation to individuals with chronic illness. Thorne *et al*., (2002) suggest the chronically ill patients who used CAM *"were taking on an increasing responsibility for their own role in disease management"* (pg. 674) and that CAM played a significant part in their self-care management and contributes to optimal wellbeing when living with a chronic disease.

An interesting overlap between Thorne's study (2002) and Brough's work (2012) which has implications for the conceptual framework is that CST/CAM users learn to 'tune into' what their bodies are telling them, and 'body listening' becomes an important part of the process of learning about, and understanding, one's body and one's self, based on any cues it provides. Being able to evaluate this process in some way is useful as it may demonstrate that patients are engaging in their own process and taking responsibility for their health and wellbeing. Items on the WHHQ that represent this process include: 'I'm learning about myself and my body', 'I've felt aware of my body's needs', 'My awareness about my health has helped me manage my life' and 'I've felt in control of my health and wellbeing'. During the process of this thesis, items such as these have not been identified in any other HRQoL measures, making them unique to health evaluation.

10.4.4 Social aspects of health

Brough's work (2012) identified that, as a result of having CST, participants had improved interpersonal relationships. Participants in the focus groups in this study

went beyond this to develop the social wellbeing domain further. As a result, the social aspects of the conceptual framework now include: use of support networks, local community and intimate relationships. The social wellbeing domain on the conceptual framework was honed to include 'connected to local community' and 'making use of a support network'. Having included these aspects into the social wellbeing domain based on CST users input, literature supporting or contradicting this decision was sought. Components of social health and functioning include social role participation, social network quality, interpersonal communication and social support (Castel *et al.*, 2008). The content validity of a prespecified conceptual model in the Patient-Reported Outcomes Measurement Information System (PROMIS) which defines social-health related outcomes as "*wellbeing regarding social activities and relationships, including the ability to relate to individuals, groups, communities, and society as a whole" (pg. 738*) was assessed in Castel's study (2008). Here, focus groups were undertaken to explore whether their conceptual framework was adequate and comprehensive. Focus groups identified volunteerism and pet ownership as important concepts and fulfilling both family and work responsibilities, with a distinction between activities done out of obligation *versus* enjoyment being made. Social role participation and satisfaction were focused on as these concepts align with outcomes rather than processes (Castel *et al*., 2008). Social role participation involves 'an engagement or taking part in' and satisfaction with usual social roles including those involving relationships with spouses or partners, children, work colleagues and leisure activities (Dikers *et al*., 2000 and McDowell and Newell, 1996 cited in Castel *et al*., 2008) were identified. Castel's results showed that participants referred to responsibilities towards other family members - especially children, in the same way that they referred to work responsibilities. Social wellbeing was often related to the degree to which they fulfilled the demands made on them for both childcare and financial security. Regarding social health role performance, the distinction between activities undertaken out of duty *versus* enjoyment, was important to participants and Castel (2008) concluded that a revision of how the concept of satisfaction was integrated into the social health domain was warranted. As regards 'role satisfaction', this was reflected on by the degree to which participants'

social role expectations, desires, needs and demands (imposed internally or externally) were fulfilled. Castel (2008) concluded that further examination of the concepts including distinctions between performance and satisfaction was needed. The conceptual framework of CST outcomes included the aspects of 'life satisfaction' and 'doing things I enjoy'. The items of the WHHQ-25 which tap these aspects of social-health are 'my daily life is full of things that keep me interested'; 'I've felt engaged in life'; 'I've felt satisfied by my school, work or current role in life' and role-performance were met by the item 'I've had too many demands made on me'. Not addressed by Castel's study, is how participants engage with support networks. In this thesis the item 'I've asked for help when I've needed it' covers that component. The component 'Making use of a support network' may be linked to taking more responsibility for self; as individuals realise what their needs are and make new choices about the support they need at different times.

10.4.5 MYCaW coding guidelines

Whilst the MYCaW (Paterson *et al.*, 2007) questionnaire was excluded based on the fact that it was cancer specific, the content of the coding guidelines for qualitative data (Polley *et al*., 2007) was reviewed to note what categories were included. Super-categories have been identified and a breakdown of what comprises each super-category is presented. On the MYCaW form at baseline, data relating to 'concerns and problems', as one would expect, are primarily symptom based and negatively positioned. Categories include: S1 psychological and emotional concerns, S2: physical concerns, S3: Hospital cancer treatment concerns, S4: concerns about wellbeing and S5: practical concerns. In the category of S4 – concerns about wellbeing, references to wider issues around living well, regaining and maintaining health are included. Explanations suggest references regarding general wellbeing such as 'taking more time for self', and references related to spiritual wellbeing: meaning and peace include improving the connection between the mind, body and/or the spirit as a means of becoming more in touch with one's self or adopting a more holistic approach to life, match the findings of this study.

At follow-up, data related to 'other things affecting your health' are categorised into 7 categories (OT1 – OT7) and are either positive or negative. The categories include OT1 'awareness of wellbeing', OT2 'receiving complementary therapies', OT3 'major life events', OT4 'social support', OT5 'work situation', OT6 'health issues', and OT7 'other'. The findings of this study mirror the explanations in category OT4: social support, which reference engaging in support from family/friends and the wider community.

10.5 The contribution this work offers

The contribution of the WHHQ to the existing literature depends to some extent on the conceptual framework and the extent to which it encompasses aspects of health and wellbeing not present in other conceptual frameworks.

So, it is interesting to compare the conceptual framework of CST outcomes with that presented in the IN-CAM database framework (Verhoef *et al.*, 2005b; Verhoef *et al.*, 2006) of outcome domains important to CAM research. The IN-CAM framework takes a global look at the domains relevant to CAM (IN-CAM, 2006) and provides the structure to the IN-CAM database. There are many similarities in the content of this conceptual framework and the CST conceptual framework developed for this thesis (Table 49). Both acknowledge the inherent overlap within domains. The two main differences are that the conceptual framework of CST outcomes is structured differently and was evaluated by CST users. Throughout this process many of the topics were discussed in terms of relevance to users, of content making sense to users, regarding experience of receiving CST and how framework contents were operationalised or related to each other. For example: on the IN-CAM framework items in the psychological domain like patient satisfaction, patient's perceptions of care, perceptions of risk, preference for control and patient expectations are more to do with the overall experience of care than with the treatment effects. The IN-CAM social domain includes economic and socio-economic items. During the focus group discussions in this thesis, no consensus was reached as to whether financial implications had a place on the conceptual framework as participants had mixed opinions about its relevance. The

inclusion of spirituality and religion was debated amongst focus groups participants. Spirituality was agreed upon and was endorsed by participants, whilst on the IN-CAM framework religiosity is identified as a component of social wellbeing and spirituality as a component of spiritual wellbeing. Five items included in the IN-CAM framework that do not feature in the CST conceptual framework are absorption, introversion, readiness, sense of coherence and transformation. Items such as biomarkers and pathology were discussed, but they were not deemed as outcomes *per se* in this context. Similarly, the CST conceptual framework intentionally did not include the process of healing or the context of healing as they, too, were not considered or categorised as outcomes in this thesis. As the conceptual framework for CST was developed, the item pool was elicited meaning there were multiple items tapping the concepts on the framework and, subsequently, the items were evaluated with CST users in cognitive interviews. The IN-CAM framework falls short of this process as its intention was to identify outcomes of importance to CAM to enable the selection or development of outcome measures suitable to assess whether CAM interventions work. Its content was not evidenced by CAM users.

Table 49: Comparison of content of CST conceptual framework v IN-CAM framework

Conceptual framework of CST outcomes **main concept:** Health and wellbeing			**IN-CAM** content taken from (Verhoef *et al.*, 2006)
Domain	**Sub domain**	**Components**	
Mental wellbeing	Responsibility for Self	Autonomy	Absorption*
		Allowing a nurturing experience	Enablement**, locus of control**
		Taking reflective downtime	
		Sense of wholeness, self-confidence and balance	Self-esteem
	Self-concept	Connection to family history and past	
		Different feelings state	Anger**, Anxiety**, depression**, empathy**, mood**, relaxation**, stress**, hope, trust, introversion*, resilience**
		Understanding and accepting emotions	
	Emotions and feelings	Understanding Mind-body-spirit links	Coping, attitudes and beliefs, optimism/pessimism**, openness to experiences**, readiness*, sense of coherence*, awareness**
		Adopting new attitudes or coping strategies	
	Mental outlook		Patients perceived knowledge‡, motivation**, self-efficacy**, patient satisfaction‡, perception of care‡, perception of risk‡, preference for control‡, patient expectations‡.

*Items marked * are not covered on the conceptual framework of CST outcomes.*
Items marked ‡ were discussed in focus groups and were not included based on CST users input and feedback.
*Items marked ** are included in components within concepts on conceptual framework of CST outcomes i.e. symptoms (pain, fatigue*

Conceptual framework of CST outcomes **main concept:** Health and wellbeing			**IN-CAM** content taken from (Verhoef *et al.*, 2006)
Domain	Sub domain	Components	
Physical wellbeing	Mobility, posture, function Daily tasks	In tune with one's body, able to read body's signals Sleep quality Independence Limitations Life satisfaction Doing things, I enjoy	Biological markers ‡ Disability ** Sleep Energy Fatigue ** Function/activities of daily living Pain** Pathology ‡ Symptom management ‡

Conceptual framework of CST outcomes **main concept:** Health and wellbeing#			**IN-CAM** content taken from (Verhoef *et al.*, 2006)
Domain	Sub domain	Components	
Spiritual wellbeing	Being present	Reflecting on meaning of life Connection to self, divine, others, nature and animals Living my core values Trust, faith in life, having compassion	Awareness Balance ** Energy Harmony ** Hope ** Peace ** Relaxation Spirituality ‡ Transformation *

Conceptual framework of CST outcomes **main concept:** Health and wellbeing			**IN-CAM** content taken from (Verhoef *et al.*, 2006)
Domain	Sub domain	Components	
Social wellbeing	Engaging in life	Intimate relationships Making use of a support network Feeling connected to family and friends Feeling connected to local community	Adjustment (participation) Advocacy ‡ Economic (Health care utilization, cost effectiveness) ‡ Relationships Role function in daily life and work Social support Socioeconomic ‡ Social strain ‡ Religiosity ‡

Conceptual framework of CST outcomes **main concept:** Health and wellbeing#		**IN-CAM** content taken from (Verhoef *et al.*, 2006)
Domain	component	
Symptoms (can manifest in any domain)	Severity, frequency Recovered Relief, temporary and sustained	Some symptoms listed in above domains (pain, fatigue)
Energy (is required for change in all domains and its components)		Shown within domain of spiritual and physical.

*Items marked * are not covered on the conceptual framework of CST outcomes.*
Items marked ‡ were discussed in focus groups and were not included based on CST users input and feedback.
*Items marked ** are included in components within concepts on conceptual framework of CST outcomes i.e. symptoms (pain, fatigue)*

Based on FDA guidance (2009) having a conceptual framework validated by patients is now the basis for PRO development. In general, PRO development has gone through important methodological shifts in the last decade; from questionnaires being developed by clinicians when clinical intuition and experiences were regarded as sufficient requirements for the construction of PRO questionnaires (Gorecki *et al.*, 2013), to ensuring the input of patients as part of the development and validation process; resulting in PROs that have more meaning to the populations in which they are used. A further transition is required for policy makers and PRO developers to recognise the importance that wellbeing has in people's lives, thus aiding a shift from the bio-medical model perspective of health care evaluation. The persistence of the medical model may be hindering the advancement of measurement methods for evaluating concepts such as wellbeing and CAM outcomes.

The contribution made through this thesis is that the components which make up the domains of conceptual framework of CST outcomes represent processes, actions or activities of how individuals relate to those aspects of their lives. In turn, the items on the questionnaire were developed to capture the majority of these outcomes. This study assumes that people's health improves after CST and, as there is no order or hierarchy in the way outcomes manifest, the arrows in the framework illustration proved to be a good way to show some of the important relationships between the domains, sub-domains and components.

The conceptual framework was not dominated by one aspect of health over another (e.g. many disease measures are focused solely around pain and physical functioning) as this seems like a balanced way of presenting outcomes of holistic health. So, whilst everyone will experience changes in the domains in a unique way, the conceptual framework represents the most common and important of these relationships as endorsed by CST users and CST practitioners in this study.

10.6 Strengths of this work

The strengths of this work lie in the fact that the conceptual framework developed during this study is the first of its kind for CST. CST users and CST practitioners were involved in all stages of the development and evaluation process of the 25-item WHHQ, enhancing and endorsing content and face validity. This validity has been evidenced as requests to use the questionnaire have been made by both CST practitioners and other practitioners in the field of CAM. An evaluation of the measurement properties of the 25-item WHHQ (pg. 151) showed that the new PRO is psychometrically sound and has good internal consistency and convergent validity with the WEMWBS and the HEHIQ. An anchor-based method to establish test-re-test reliability also showed good results: errors were calculated as similar to the WEMWBS and SF12. Respondents reporting improvements in their HRQoL, showed small changes above this threshold during the evaluation of responsiveness.

The second strength is that the conceptual framework identified important domains of health affected by CST that, until now, have not been explicitly expressed as statements or items on any PROs, even those created for CAM. These include the development of self-awareness and individuals taking responsibility for themselves.

The third strength is that this study enabled the development and testing of a PRO that widens the meaning of health and wellbeing. This is because items have been included that tap the concept of self-awareness such as 'my awareness about my health has helped me manage my life' and 'I'm learning about myself and my body' and the concept of individuals taking responsibility for self, 'I've felt in control of my health and wellbeing'. These items are novel in healthcare evaluation as all require patients participating in the management of their own health.

The fourth strength is that because the 25-item WHHQ is based on a detailed qualitative examination of patient perceived outcomes, the WHHQ can capture the full range of experiences, enabling individuals to reflect on change in aspects of their health and wellbeing they might not have thought would change. THE WHHQ

also offers an opportunity to expand the client's knowledge of health and wellbeing by changing the mindset of the individual completing the questionnaire.

The relationship between the researcher and CST practitioners has enabled much dialogue, which also strengthened the study. The researcher listened to CST practitioner concerns about barriers to using the questionnaire, including those working from home who did not have a waiting area; meaning that asking their clients to complete the questionnaire would take up valuable session time and practitioners expressed their concern that their presence might influence client responses or that they might be asked for advice regarding completion.

Some practitioners suggested that, as a result of having had clients complete the WHHQ, discussions were initiated which may have not arisen without the prompt, adding value to the therapeutic relationship. In addition, a conversation about CST research in the wider context, such as clinical audit and building an evidence base has begun.

10.7 Limitations of this work

The main limitation of this work was that the sample comprised of a self-selected group who were keen to support CST research, introducing the possibility of bias. People with negative experiences of CST or those struggling with symptoms may have not volunteered to participate in this study. Most of the sample were female introducing another possible source of bias, as men may have a different experience of CST than women. This was not evident from the small number of men who were involved in this study but does not rule out the possibility.

Another limitation was that the sample sizes both in the qualitative and quantitative aspects of this study were small. The sample sizes in the sub-studies to assess measurement properties of the 52-item WHHQ and then to evaluate the measurement properties of the 25-item WHHQ were smaller than the literature recommends (Streiner *et al.*, 2015) and, therefore, results should be interpreted with caution. Whilst there are no clear recommendations about the number of

subject to item ratios for carrying out EFA (Anthoine *et al.*, 2014) the communalities in this thesis will have been influenced by the sample size.

The recruitment process in this thesis had multiple layers due to the need to engage both CST users and CST practitioners in the process of developing and evaluating the conceptual framework and the WHHQ. Whilst efforts to attract CST practitioners who were not members of the CSTA were undertaken *via* telephone calls and emails to key research individuals within other member organisations of CST, only two non-member practitioners took part in the research process. It is difficult to say how this may, or may not, have influenced the outcomes. Engaging CST practitioners in the recruitment of participants meant that the sample had a range of age and gender characteristics and diverse health problems. The sample also had differing degrees of experiences in the length of time that they had been receiving CST. However, only 8% of participants in sample two were having CST for the first time and this may be why the 25-item WHHQ results only show a small amount of change, as the majority of CST users were having maintenance sessions and considered themselves to be in good health.

10.7.1 Ethical Tensions of being a Practitioner/Researcher

Ethical tensions were also present due to NB being a practitioner researcher. Role conflicts had to be managed and the biases associated with this were considered during the design phase of this study and reflexively throughout. This was mitigated through supervision by academics that have no allegiance to the field of CST. Although SSB is known to be a CAM researcher, the research was also scrutinized by a second supervisor with no such links to CAM (HP).

The decision to undertake this research as a PhD candidate/student has had its benefits such as: researching a topic which is close to my heart, receiving a high level of training and obtaining experience in research methods. Furthermore, as a consequence of completing this study, I may be perceived as an 'expert' in the field as no one else has undertaken work in this area to date. My professional profile amongst the CST community has risen due to the presentations I have given,

publications I have written and the networking I have undertaken. However, balancing the roles of CST practitioner and researcher brought rise to personal tensions throughout the course of the study. Being a CST practitioner helped when contacting other practitioners, especially as the support of the CSTA gave the use of the membership as a means of promoting the study, potentially engaging both practitioners and to gain access to their clients as potential participants. I felt like I was accepted by practitioners because I had experience of CST.

When engaging with practitioners at different times throughout the study it was important to build rapport and maintain good relationships as the practitioners were 'gate keepers' to their clients. However, to say that I had this as an intention beforehand feels like I would be saying that I had ulterior motives for building those relationships in that way, but I genuinely have an interest in people/practitioners and how they feel about evaluating what they do. This research may not have happened without my practitioner experience.

These processes have been very integrated, this may be since the CSTA funded the research activities and this commitment allowed members to be engaged in the progression of the study. Having given practitioners permission to gather data and evaluate their work has also raised the awareness around the subject of patient reported outcomes and healthcare evaluation, enabled discussions about what areas CST supports in the lives of users and revealed that practitioners are keen to implement the WHHQ in their practices.

However, the disadvantages came in the form of having to meet contractual obligations with different stakeholders which included terms that were not implicit. For example, I did not set up financial arrangements between the university and the CSTA. Thus, there was no formal contract between the two main stakeholders in place prior to the PhD and no agreement regarding copyright or ownership of the WHHQ. Had I have been aware of, or informed about, matters such as these when enrolling for my PhD the route to implementation may have been easier. Conflicting expectations from those stakeholders included: having to complete the study to rigorous research standards, and to provide a 'tool' quickly for use by the

CST community for implementation. All of which added additional pressure to a process which is inherently stressful.

10.8 Questionnaire implementation

The intended use was research and evaluation in clinical practice. It is possible to reflect with slightly different goals, for example: if developing the WHHQ primarily for research purposes but it will, however, be invaluable for research of CST and other CAM modalities which capture these outcomes. The implementation of the WHHQ will require further work and decisions about how to manage this need to be made. Coming to an equitable and reasonable agreement around licensing is important. Anyone wanting to use the questionnaire will need to register to use it. This provides an opportunity to gather data on the measure as people register, which allows for monitoring spread and impact:

a) so that the author knows who is using the instrument (e.g. CST practitioners, researchers, non-CST practitioners, other healthcare practitioners)
b) for what purpose the instrument is being used (e.g. as expected during design, or for novel use) and;
c) where (e.g. in the UK or elsewhere) and in what settings (e.g. clinical, research) the questionnaire is being used.

Acknowledgement will be made to the authors, funders and copyright holders. To maintain the integrity of the questionnaire a copyright notice will be enforced and will include the boundaries of use. For example:

"The Warwick Holistic Health Questionnaire is subject to copyright. All rights reserved. Any unauthorised copying, changing of the wording, layout or content of the questionnaire will constitute an infringement of copyright".

Dissemination of information about the WHHQ will be required by the author and interested parties, informing people of its existence, how to use and how to register for use. In a field like CST where people are not routinely trained in research, they may need help in interpreting results.

At the time of writing this thesis, deliberations are still underway about whether to charge for use and, if so, for which groups, for example: CST users outside of the UK and other CAM professions. The management of translations would require separate agreements and consideration. However, such projects are outside the scope of this thesis.

10.8.1 Managing the implementation of the WHHQ for the CSTA

At the time of writing this thesis, the ways in which the WHHQ may be made available for use are under discussion, though a common practice within the field of patient reported outcome measures is *via* a licensing model. A license is a business arrangement in which one organisation gives another organisation permission for use under certain terms. An example of a potential scenario for the CSTA that will be investigated in over the coming months is discussed next.

The CSTA, who part funded this study, may act as a gatekeeper for any CSTA practitioners to be able to access the WHHQ via their membership status. This will enable practitioners to: a) register to use the WHHQ b) have access to a user manual and basic training on the practicalities of how to use the questionnaire for data collection within their clinic c) get support with data input *via* online resources d) be able to send collected data to a central point with the CSTA, contributing to a wider pool of data e) be able to access guidance on how to interpret their results.

10.9 Future research

Scale development is an iterative process as the FDA guidance demonstrates and so the instrument will continue to evolve. For example, as the WHHQ is used in different samples, the estimates of validity and reliability may change and, therefore, refinements to the measure may be needed.

Analysis of the collaborative data collection from the CSTA practitioner members will be undertaken and reported on in an appropriate way. The data collected within the first 12-month period may also be used for further validation work on the WHHQ. An implementation phase such as this will be closely monitored and queries handled by the author.

10.9.1 Refining the 25-item WHHQ

- Test-retest to be revisited using larger sample.

- To build on the validation work presented in this thesis, the WHHQ may be evaluated using other methods. A larger sample of data could be collected enabling a confirmatory factor analysis (CFA). CFA is used to examine the nature of and relations amongst latent constructs (Jackson *et al.*, 2009).

- Further testing is needed to establish the WHHQ's capacity to detect true change accurately if the WHHQ is to be used to evaluate CST as an intervention. Undertaking a longitudinal study would be one way of doing this and it will also be important to ensure that the sample includes many new clients and also some whose condition has deteriorated during treatment.

- Further research testing the 25-item WHHQ in its current form (80% positive items and 20% negative items) and comparing it against a 100% positive 25-item WHHQ version to determine the best way of limiting response bias.

10.9.2 Research setting

Once the validity of the above is satisfied the 25-item WHHQ could be used for intervention and evaluation research to establish the effectiveness of CST as an intervention to enhance health and wellbeing.

10.9.3 Evaluating clinical practice

Future work to establish evaluation as a normal part of CST practice *via* CST associations and their member practitioners is envisioned. It is important to evaluate the extent to which WHHQ can enhance the therapeutic encounter, thus improving communication, prompting discussion and monitoring changes and outcomes of CST sessions.

10.9.4 Electronic version of the 25-item WHHQ

An electronic version of the questionnaire which undertakes scoring and reports practitioner based outcomes collectively is currently being appraised. It is anticipated that, if an electronic version is acceptable, this may eliminate such barriers for use and may be useful for practitioners working from home or those who wish to have more flexibility for their clients.

10.9.5 Use in other populations

The 25-item WHHQ may be suitable for use in other CAM modalities and to evidence this, evaluation to the establish validity, reliability and responsiveness to change needs to be carried out in different populations. For example, with individuals having reflexology.

11 Conclusions

None of the PROs reviewed in this thesis had provision for all of the outcomes important to CST users. In those that were conceptually a good match, issues with poor reliability and validity were present. Trials of CST to date have used *ad-hoc* questionnaires which fail to capture all outcomes of relevance. The development of a new PRO specifically for CST was appropriate.

Identifying conceptual domains important to CST users provided a useful framework to embark on the development of the instrument. The conceptual framework developed was based on the assessment of treatment benefit as perceived by users of CST.

The WHHQ was found to be acceptable and reliable having good internal consistency, face and content validity in line with the FDA criteria. Reporting in this thesis has adhered to the COSMIN checklist when possible, supporting the WHHQ's use with CST users in clinical practice and future research.

Although more validation work is required, these results are sufficient to recommend the WHHQ for use in clinical settings at present to enable practitioners to establish the process of clinical evaluation. This process could be used to generate data for further studies to refine the measure.

There are potential concerns with the WHHQ's responsiveness which may be due to the limitations of the sample in which responsiveness was tested. Further testing is needed to determine the WHHQ's ability to detect true change accurately if the WHHQ is to be used to evaluate CST as an intervention.

This study has made an important contribution to the clinical practice of CST. The mixed methods approach embedding the perspectives of CST users and CST practitioners throughout the development and validation process worked well and enabled practitioners to define the outcomes of their practice in a way that has not been done systematically before.

The 25-item WHHQ thus provides a useful tool to enable clinical evaluation. It has the potential to measure the changes in the health and wellbeing of people having CST and to enable more robust studies of the effectiveness of CST to be carried out. In the future, such studies could influence provision of CST in health services.

12 Appendices

12.1 Appendix 1: Ethics approval letter

20th May 2015

Warwick Medical School

PRIVATE
Miss Nicola Brough

Dear Miss Brough,

Study Title and BSREC Reference: *Designing and evaluating a health related quality of life outcome measure for craniosacral therapy (CST)?* REGO-2015-1499

Thank you for submitting the above-named project to the University of Warwick Biomedical and Scientific Research Ethics Committee for research ethical review.

I am pleased to advise that research ethical approval is granted.

May I take this opportunity to wish you success with the study, and to remind you that any substantial amendments require approval from BSREC before they can be implemented. Please keep a copy of the original signed version of this letter with your study documentation.

Yours sincerely

[redacted]

Professor Scott Weich
Chair
Biomedical and Scientific
Research Ethics Sub-Committee

Biomedical and Scientific
Research Ethics Sub-Committee
A010 Medical School Building
Warwick Medical School,
Coventry, CV4 7AL.
Tel: 02476-528207
Email: BSREC@Warwick.ac.uk

Medical School Building
The University of Warwick
Coventry CV4 7AL United Kingdom
Tel: +44 (0)24 7657 4880
Fax: +44 (0)24 7652 8375

THE UNIVERSITY OF WARWICK

12.2 Appendix 2: Recruitment poster

Having Craniosacral Therapy (CST)? Can we measure outcomes that are important to you appropriately?

Would you like to take part in this research project?

We would really value your input in our current research project. The aim of the research is to design and evaluate a questionnaire that can assess changes in health and wellbeing of people who choose to have CST.

Participants
We are looking for individuals who have had at least four sessions of CST in the past, or are currently having CST.

What does participation involve?

- Giving your consent.
- To evaluate a conceptual framework describing the possible effects of treatment with CST, taking part in an audio recorded focus group (of up to 8 people) to discuss your thoughts. Comments will be used to improve the conceptual framework which will be used to generate the items in the draft questionnaire (to be held **on Saturday 18th July 2015** in a location convenient to participants when possible).
- Or, to take part in a focus group to assess the draft questionnaire design, layout and instructions and to reflect on the comprehensibility of each item, response format and time frames (to be held in **Headington, Oxford, August 2015** date to be arranged).
- Or, to undertake an audio recorded interview (at a time and place convenient to you) to complete a questionnaire to discuss any queries you may have in understanding the items or answers and any challenges presented in doing so (**to be conducted in September 2015**).
- Or, to complete the questionnaire(s) before and after having CST e.g. before 1st and 3rd sessions (**January 2016 - March 2016**) and (**August 2016 - October 2016**).

Why is research important?
Research evidence improves the credibility of an intervention like CST and increases the possibility of CST being used alongside conventional medicine (for use within or primary or secondary care within the National Health Service) to meet better patient outcomes.

Background
Public interest in and engagement with CAM are rising, yet there are concerns about the evidence base for different therapies; it is in response to this that the motivation for this research has arisen. In order to assess the effectiveness of craniosacral therapy a suitable questionnaire assessing health related quality of life and outcomes that are important to users is required. The CSTA are sponsoring this study in conjunction with the University of Warwick (Faculty of Medicine).

What to do next - Having read the outline above, should you wish to participate in this study or seek further information please contact your local CST practitioner or contact: Nicola Brough on 07960946853 or email: N.Brough@warwick.ac.uk. (All focus groups will be held in a location most convenient to participants, travel expenses will be met up to £50).

12.3 Appendix 3: Participant information leaflet

Study Title:	Designing and evaluating a health-related quality of life outcome measure for Craniosacral therapy (CST)?
Investigator(s):	Sarah Stewart-Brown (Chief Investigator), Helen Parsons (Academic Supervisor), Nicola Brough (Phd Candidate).

Introduction

You are invited to take part in a research study. Before you decide, you need to understand why the research is being done and what it would involve for you. Please take the time to read the following information carefully. Talk to others about the study if you wish.

(Part 1 tells you the purpose of the study and what will happen to you if you take part. Part 2 gives you more detailed information about the conduct of the study)

Please ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

PART 1

Do I have to take part?

This is entirely up to you. I will describe the study and go through this information sheet, which you can keep. If you choose to participate in the focus groups or interviews we will ask you to sign a consent form to confirm that you have agreed to take part. If you chose take part in steps 5 & 6 you will be deemed to have consented, by returning a completed questionnaire.

You will be free to withdraw from any stage of the research at any time, without giving a reason and this will not affect you or the care you receive from your CST practitioner.

What will happen to me if I take part?

As a participant in step 2.

- You will be asked to reflect on the conceptual framework with the help of a questionnaire which you will bring along to the focus group (up to 6 people) for discussion.

- You be asked to share your thoughts and ideas with the group.
- The discussion will take between 60 - 120 minutes and will be recorded.
- This group will be held on Saturday 18th July.2015 in a place convenient to participants when possible.

As a participant in step 3.

- You will be given a draft questionnaire to complete and reflect on.
- You in be invited to share your thoughts and ideas about questionnaire's design, layout and instructions and to reflect on the comprehensibility of each item, the appropriateness of the answers and the time frames in a focus group (up to 6 people).
- The discussion will take between 60 – 120 minutes and will be video recorded.
- This group is for clients and will take place in Edinburgh City Centre (Date to be confirmed September 2015).

As a participant in step 4.

- You will take part in a one-to-one interview at a time and place convenient to you. The place needs to be suitable for recording purposes.
- In the interview, you will be asked to complete a questionnaire and to discuss any queries you may have in understanding the items or answers and any challenges presented in doing so.
- The interview will be recorded and will take the necessary time required to address the points raised.

As a participant in step 5.

- You will be asked to complete a questionnaire after a session of CST. For example: at the beginning of your second session.

As a participant in step 6.

- You will be asked to complete four different questionnaires before and after a session of CST. This will allow us to compare the new questionnaire with others.
-

What are the possible disadvantages, side effects, risks, and/or discomforts of taking part in this study?

You are free to disclose whatever information you choose about your experiences of CST and are not required to disclose information of a sensitive, embarrassing or upsetting nature.

What are the possible benefits of taking part in this study?

There are no specific benefits to you for taking part in this study. However, the outcome measure will have benefits for others.

Expenses and payments

You will not be reimbursed for time spent on this research although light refreshments will be provided in focus groups and interviews. You can be reimbursed all of your travel expenses if the full amount is less than £50 or can claim up to £50 for travel expenses over this amount, on producing valid travel tickets (train & bus), parking tickets and mileage can be claimed at 45p per mile up to the value of the amount stated. On accepting the receipts, the researcher (Nicola Brough) will ensure the expenses are reimbursed within 2 weeks of that date by BACs transfer, via the CSTA Treasurer, Mr Vincent Winter.

What will happen when the study ends?

Any data I collect will be stored for 12 months. At that point any data on paper will be shredded and disposed of in secure disposal facilities. Electronic data will be transferred to non-rewritable cds and placed with all other study documentation in the archives within the University.

Will my taking part be kept confidential?

Yes. We will follow strict ethical and legal practice and all information about you will be handled in confidence. Further details are included in Part 2.

What if there is a problem?

Any complaint about the way you have been dealt with during the study or any possible harm that you might suffer will be addressed. Detailed information is given in Part 2.

This concludes Part 1.

If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.

PART 2

Who is organising and funding the study?

Nicola Brough is organising this project and the research expenses are being funded by the Craniosacral Therapy Association (UK) and a Chancellor's Scholarship Award. Nicola is a PhD student supervised by Prof Sarah Stewart-Brown and Dr Helen Parsons. This study is being conducted as part of a PhD in the department of Health Sciences, Warwick Medical School.

What will happen if I don't want to carry on being part of the study?

Participation in this study is entirely voluntary. Refusal to participate will not affect you in any way. If you decide to take part in the study, you will need to sign a consent form, which states that you have given your consent to participate.

If you agree to participate, you may nevertheless withdraw from the study at any time without affecting you in any way.

You have the right to withdraw from the study completely and decline any further contact by study staff after you withdraw. Withdrawal from the study will not affect the relationship you have or usual care from your CST practitioner.

What if there is a problem?

This study is covered by the University of Warwick's insurance and indemnity cover. If you have an issue, please contact Nicola Brough (Researcher) or Professor Sarah Stewart-Brown (Chief Investigator) whose details can be found at the end of this information leaflet.

Who should I contact if I wish to make a complaint?

Any complaint about the way you have been dealt with during the study or any possible harm you might have suffered will be addressed. Please address your complaint to the person below, who is a senior University of Warwick official entirely independent of this study:

Director of Delivery Assurance
Registrar's Office
University House
University of Warwick
Coventry
CV4 8UW
Complaints@Warwick.ac.uk
024 7657 4774

Will my taking part be kept confidential?

In order to ensure confidentiality, the data protection act (1998) will be adhered to and the anonymisation or pseudonymisation of data will be undertaken. Electronic data will be backed up to a external hard drive and stored securely. Only the research team including the academic supervisors and researcher will have access to the data. All data held on paper will be kept under lock and key with only the researcher having access to it directly. Names and interview data will be stored separately. Should someone other than the researcher be chosen to transcribe the focus group discussion, they will be asked to sign an agreement to observe confidentiality of the data.

What will happen to the results of the study?

This study is expected to last for 2 years. At the end of the study we will publish the findings in a peer reviewed journal and a full report will be available on the CSTA website (accessible to members only) and may be discussed at conferences. If you would like to obtain a copy of the published results, please ask your CST practitioner.

Who has reviewed the study?

This study has been reviewed and given favourable opinion by the University of Warwick's Biomedical and Scientific Research Ethics Committee (BSREC): **REGO-2015-1499. 20th May 2015.**

What if I want more information about the study?

If you have any questions about any aspect of the study or your participation in it, that is not answered by this participant information leaflet, please contact:

Nicola Brough, email: n.brough@warwick.ac.uk and contact: 07960 946853

Nicola's Academic Supervisor, Sarah Stewart-Brown, email: sarah.stewart-brown@warwick.ac.uk and contact: 0247 6574510.

Thank you for taking the time to read this participant information leaflet.

12.4 Appendix 4: Revised working conceptual framework of CST outcomes

Date	Revisions made to working conceptual framework of CST outcomes
6 January 2015	Developed 'CST intervention HRQL Conceptual Framework' using thematic map of clients' experiences from previous work & literature.
28 January 2015	Based on discussion with supervisors the following changes were made: • document title changed to Conceptual Framework for CST v2 • social domain included • sub domain of embodiment and components included under physical domain • domains of relating and awareness removed and circles representing the process of relating and awareness included, with hypothesis text inserted to inner circle of diagram.
09 February 2015	• document title change to v3 • sub domain-symptoms and components were moved to right side to connect with physical domain • Main construct changed from CST intervention HRQL to wellbeing
03 March 2015	To aid understanding a list of definitions of key concepts for the working framework of CST outcomes was developed. For use in conjunction with this document.
08 March 2015	• Inserted word 'health' in key construct to read 'health and wellbeing' to match definitions.
18 March 2015	• Based on PhD upgrade panel feedback title changed to 'working' conceptual framework of CST outcomes v4.

13 April 2015	Based on discussion with supervisors the following changes were made: • Construct labels changed: psychological wellbeing to mental wellbeing, social to social wellbeing. • Sub domain labels changed: cognitive to mental outlook; affective to emotions/feelings. • Sub domain: embodiment removed and component: in tune with my body; able to read body's signals linked to sub domain, Mobility, posture, function. • Inserted arrows from domain: physical functioning to relevant sub domains • Changed colours to assorted fills (for grey scale printing) and inserted fill type in Key. Document title changed to v5.
June 2015	V5 was sent to participants for evaluation.

12.5 Appendix 5: List of draft items (73-WHHQ)

List of draft Items for interviews

Below are some statements about your health and wellbeing. Please tick the box that best describes your experience over the last two weeks. Please make a selection for all statements.

STATEMENTS

	Physical functioning	None of the time	Rarely	Sometimes	Often	All of the time
1	I've been physically independent	0	1	2	3	4
2	I've been able to read my body's signals	0	1	2	3	4
3	I've had lots of energy	0	1	2	3	4
4	I've been sleeping well	0	1	2	3	4
5	I've been in tune with my body	0	1	2	3	4
	Symptoms	None of the time	Rarely	Sometimes	Often	All of the time
6	I've been in pain	0	1	2	3	4
7	My symptoms have been a problem	0	1	2	3	4
8R	I've had relief from my symptoms	0	1	2	3	4
9	My health limits my daily activities	0	1	2	3	4
	Everyday life	None of the time	Rarely	Sometimes	Often	All of the time
10	I feel satisfied by my school, work or current role in life	0	1	2	3	4
11	I feel satisfied with my work/life balance	0	1	2	3	4
12	I feel satisfied with my life overall	0	1	2	3	4
13	I have too many demands made on me	0	1	2	3	4
14	I've engaged in nurturing activities	0	1	2	3	4
15	My daily life is full of things that keep me interested	0	1	2	3	4
	Mental wellbeing/mental outlook	Disagree strongly	Disagree somewhat	Don't know	Agree somewhat	Agree strongly
16	I live life one day at a time and don't really think about the future	0	1	2	3	4
17	I've been able to adapt to meet life's challenges	0	1	2	3	4

18	Life has been a continuous process of learning, changing and growth	0	1	2	3	4
19	I've been feeling positive about life	0	1	2	3	4
20	I've been coping with daily life	0	1	2	3	4
21	I believe my health is linked to my daily life	0	1	2	3	4
22	The perspective I have about my condition helps me manage life	0	1	2	3	4
23	I tend to bounce back quickly after hard times	0	1	2	3	4
	Self-concept	None of the time	Rarely	Sometimes	Often	All of the time
24	I keep on learning about myself and my body	0	1	2	3	4
25	I judge myself by what I think is important, not by the values of what others thinks is important	Disagree strongly	Disagree somewhat	Don't know	Agree somewhat	Agree strongly
26	I've felt in balance	0	1	2	3	4
27	I've been feeling confident	0	1	2	3	4
	Responsibility for self	Disagree strongly	Disagree somewhat	Don't know	Agree somewhat	Agree strongly
28	Taking an active role in my own health is the most important factor in determining my wellbeing	0	1	2	3	4
29	I deal consciously with myself	0	1	2	3	4
30	I feel able to take care of myself	0	1	2	3	4
31	I've taken reflective down time	0	1	2	3	4
32	I've asked for help when I've needed it	0	1	2	3	4
33	Making sense of the past helps me figure out what to do in the present	None of the time	Rarely	Sometimes	Often	All of the time
34	I feel able to take care of myself	0	1	2	3	4
35	I've identified activities that I enjoy	0	1	2	3	4
	Emotions and feelings	None of the time	Rarely	Sometimes	Often	All of the time
36 R	I've been feeling anxious	0	1	2	3	4
37	I've been feeling confident	0	1	2	3	4
38	I've been able to express how I feel	0	1	2	3	4
39	I've been feeling calm	0	1	2	3	4
40	I've been feeling joyful	0	1	2	3	4

41 R	I've been feeling sad	0	1	2	3	4
	Spirituality	None of the time	Rarely	Sometimes	Often	All of the time
42	I've felt my inner strength	0	1	2	3	4
43	I've trusted others	0	1	2	3	4
44	I've been compassionate to myself	0	1	2	3	4
45	I've been compassionate towards others	0	1	2	3	4
46	I've had faith in life's journey	0	1	2	3	4
47	I reflect on the meaning of life	0	1	2	3	4
48	I've been feeling connected to God or the divine	0	1	2	3	4
49	I've felt connected to nature	0	1	2	3	4
50	I have a spiritual orientation in life	0	1	2	3	4
	Social Wellbeing	None of the time	Rarely	Sometimes	Often	All of the time
51	I interact consciously with others	0	1	2	3	4
52	I interact consciously with my environment	0	1	2	3	4
53	I've been feeling engaged in life	0	1	2	3	4
54	I've felt connected to my friends and family	0	1	2	3	4
55	I've identified people who help me take care of myself	0	1	2	3	4
56	I have some meaningful relationships	Disagree strongly	Disagree somewhat	Don't know	Agree somewhat	Agree strongly
	Holistic wellbeing	None of the time	Rarely	Sometimes	Often	All of the time
57	I've had a sense of wellbeing	0	1	2	3	4
58	I've been aware of the connection between mind, body and spirit	0	1	2	3	4
59	I trust my intuition	0	1	2	3	4
60	I'm self-aware	0	1	2	3	4
61	I strive for more broad awareness	0	1	2	3	4
62	I'm living authentically	0	1	2	3	4
63	I've been feeling well	0	1	2	3	4

Please continue over the pg.

To what extent do these statements reflect your recent experience?

		Does not apply	Disagree strongly	Disagree somewhat	Agree somewhat	Agree strongly
64	I feel more connected to my body	0	1	2	3	4
65	I feel more aware of the way my mind and body work together	0	1	2	3	4
66	I feel more aware of the way my mind and body are affected by my physical environment	0	1	2	3	4
67	I feel more aware of the way my mind and body are affected by my social environment	0	1	2	3	4
68	I am more aware of what my body needs to stay healthy	0	1	2	3	4
69	I can read my body signals better	0	1	2	3	4
70	I can trust my body to tell me what it needs	0	1	2	3	4
71	I feel less rejecting of my body and its symptoms	0	1	2	3	4
72	I feel less ashamed of the way my body is currently working	0	1	2	3	4
73	I feel less angry at my situation now	0	1	2	3	4

Please check you have answered all of the questions. Thank you for taking the time to complete this questionnaire.

12.6 Appendix 6: Interview schedule for cognitive interviews

Opening

A. (**Establish Rapport**) [shake hands] My name is Nicola Brough and I am the researcher for this study.

[I will show my university identity badge. And notify the participant that I will be leaving my mobile phone on during the interview, should my supervisor need to contact me. I will then ask permission to set up the recording equipment.]

Thank you for agreeing to take part in the pre-testing of the CST questionnaire. Thank you for completing the consent form, I would like to remind you that I will be recording the session, do you still consent to this? If at any point during the interview you wish to stop the process, you can say stop. If you need a comfort break that is ok too. I will notify you once I am about to start recording.

B. (**Purpose**) The aim of the interview is for me find out if the questionnaire is user friendly and easy to understand. This is done by asking you to complete the questionnaire, I will remain silent whilst you are undertaking this task, unless any issues in completion or understanding arise with regards to the items, response options or format. At which point I will probe for more information regarding the issue at hand.

C. (**Motivation**) This will allow me to evaluate how easy it is for you to complete the questionnaire and if the language, format and response options used are acceptable.

D. (**Time Line)** The questionnaire should take about 5-8 minutes to complete, and in addition we will have the opportunity to discuss any challenges if and when they arise. The whole session may take between 45-60 minutes, is this ok for you?

(**Transition:** Let me begin by asking you some questions about yourself, what's your date of birth? How many sessions of CST have you had? How frequently do you have CST?)

Body

[Present the questionnaire on a clip board and provide a pen] Please take your time and complete the questionnaire, should you have any questions, if something is unclear or you do not understand something please ask. This is not a test in your ability but an evaluation of the questionnaire itself, so any feedback will be useful for the purpose of this evaluation.

(Topic) Instructions - how did you feel about the instructions?

(Topic) Format - what did you like best about the format?

(Topic) Response options - what are the pros and cons of these response options? what would you suggest here, instead?

(Topic) else Items - what is it about that item that you don't understand? How could it be phrased? [make note of item number]

Probes for clarity:

what do you mean....

Could you tell me a little more about that ...

Would you explain further....

Could you give me an example ...

(**Transition:** Well, it has been really helpful. Let me briefly summarize the information that I have recorded during our interview.)

III **Closing**

A. (Summarize) You completed the questionnaire and the we discussed the following issues - [I will make a note here during the interview].

B. (Maintain Rapport) I appreciate the time you took for this interview. Is there anything else you think would be helpful for me to consider in relation to this questionnaire?

C. (Action to be taken) I should have all the information I need. Would it be alright to call you at home if I have any more questions? The anonymised data from this study will be published once this study is completed, did you sign to say whether or not you would like a copy of the final report? Thanks again. End.

12.7 Appendix 7: A2 Conceptual Framework of CST outcomes

Reference List

Abbot, N. C., Harkness, E. F., Stevinson, C., Marshall, F. P., Conn, D. A. & Ernst, E. (2001) Spiritual healing as a therapy for chronic pain: a randomized, clinical trial. *Pain*, 91 (1-2): 79-89.

ACQoL *Australian Centre on Quality of Life: Instruments*. [online] Available from: http:// www.deakin.edu.au/research/acqol/instruments/instrument.php (Accessed 6 March).

American Educational Research Association, A. P. A., and National Council on Measurement in Education (1999) Standards for educational and psychological testing. *American Educational Research Association, Washington DC*,

Anon - Content Validity—Establishing and Reporting the Evidence in Newly Developed Patient-Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO Good Research Practices Task Force Report: Part 2—Assessing Respondent Understanding. - *Value in Health*, (- 8): - 978.

Anon (2015) *Mental Health Today...And Tomorrow: Exploring current and future trends in mental health.* Mental Health Foundation & Pavilion Publishing and Media Limited.

Anthoine, E., Moret, L., Regnault, A., Sébille, V. & Hardouin, J. (2014) Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health and Quality of Life Outcomes.*, 12 (2):

Arnadottir, T. S. & Sigurdardottir, A. K. (2013) Is craniosacral therapy effective for migraine? Tested with HIT-6 Questionnaire. *Complementary Therapies in Clinical Practice*, 19 (1): 11-14.

Atkinson, M., Wishart, P., Wasil, B. & Robinson, J. (2004) The Self-Perception and Relationships Tool (S-PRT): A novel approach to the measurement of subjective health-related quality of life. *Health and Quality of Life Outcomes*, 2 (1): 36.

Banister, P., Burman, E., Parker, I., Taylor, M. & Tindall, C. (1994) *Qualitative Methods in Psychology: A research guide*. Buckingham, PA.: Open University Press.

Bishop, F. L. (2015) Using Mixed Methods research designs in health psychology: An illustrated discussion from a pragmatist perspective. *British Journal of Health Psychology*, 20 5-20.

Bishop, F. L., Barlow, F., Walker, J., McDermott, C. & Lewith, G. T. (2010) The Development and Validation of an outcome measure for spiritual healing: A mixed method study. *Psychotherapy and Psychosomatics*, 253 1- 13.

Bishop, F. L. & Holmes, M. M. (2013) Mixed Methods in CAM Research: A systematic Review of Studies Published in 2012. *Evidence-Based Complementary and Alternative Medicine*,

Bland, L. & Altman, D. (1986) Statistical Methods for assessing agreement between two methods of clinical measurement. *Lancet*, 307-310.

Bossuyt, P., Reitsma, J., Bruns, D., Gatsonis, C., Glasziou, P., Irwig, I. & et al. (2003) Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *BMJ*, 326 41- 44.

Braun, M., Schwickert, M., Nielsen, A., Brunnhuber, S., Dobos, G., Musial, F., Lüdtke, R. & Michalsen, A. (2011) Effectiveness of traditional Chinese “gua sha” therapy in patients with chronic neck pain: a randomized controlled trial. *Pain Medicine*, 12 (3): 362-369.

Brazier, J. E., Harper, R., Jones, N. M. B., O'Cathain, A., Thomas, K. J., Usherwood, T. & Westlake, L. (1992) Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ*, 305 (July): 160-164.

Brooks, R. (1996) EuroQol: the current state of play. *Health Policy*, 37 53-72.

Brough, N. (2012). *Craniosacral Therapy: exploring clients' experiences. A qualitative study.* Master of Philosophy in Health Science. Warwickshire, UK: University of Warwick.

Brough, N., Lindenmeyer, A., Thistlethwaite, J., Lewith, G. & Stewart-Brown, S. (2015) Perspectives on the effect and mechanisms of craniosacral therapy: A qualitative study of users' views. *European Journal of Integrative Medicine*, 7 172-183.

Brough, N., Parsons, H. & Stewart-Brown, S. (2016) The Warwick holistic health questionnaire (WHHQ), developing and evaluating a patient reported outcome measure (PROM) for craniosacral therapy (CST): a mixed methods study.

Brough, N., Stewart-Brown, S., Lindenmeyer, A., Thistlethwaite, J. & Lewith, G. (2013) Exploring clients' experiences of craniosacral therapy: A qualitative study (2012). *European Journal of Integrative Medicine*, 5 (6): 575.

Brough, N. P., H. Stewart-Brown, S. (2014) Developing and validating an outcome measure for craniosacral therapy: A mixed methods study. *European Journal of Integrative Medicine*, 6 (5): 612.

Bussing, A., Fischer, J., Haller, A., Heusser, P., Ostermann, T. & Matthiessen, P. (2009) Validation of the brief multidimensional life satisfaction scale in patients with chronic disease. *European Journal of Medicine Research*, 14 (171-177):

Bussing, A., Ostermann, T. & Matthiessen, P. (2007) Distinct expressions of vital spirituality. The ASP questionnaire as an explorative research tool. *Journal of Religion and Health*, 46 267-286.

Cardaciotto, L., Herbert, J., Forman, E., Moitra, E. & Farrow, V. (2008) The assessment of present-moment awareness and acceptance: the Philadelphia Mindfulness Scale. Assessment 15 204-223. Available from: (Accessed

Castel, L., Williams, K., Bosworth, H., Eisen, S., Hahn, A., DeWalt, D. & DeVellis, R. (2008) Content Validity in the PROMIS social-health domain: a qualitative analysis of focus-group data. *Quality of Life Research*, 17 737-749.

Castro-Sanchez, A. M., Lara-Palomo, I. C., Mataran-Penarrocha, G. A., Saavedra-Hernandez, M., Perez-Marmol, J. M. & Aguilar-Ferrandiz, M. E. (2016) Benefits of Craniosacral Therapy in Patients with Chronic Low Back Pain: A Randomized Controlled Trial. *Journal of Alternative & Complementary Medicine*, 22 (8): 650-657.

Centre, N.-H. a. S. C. I. (2013) *Provisional Monthly Patient Reported Outcome Measures (PROMs) in England - PROMS_Guide_v5.pdf*. Available from: http://www.hscic.gov.uk/media/1537/A-Guide-to-PROMs-Methodology/pdf/PROMS_Guide_v5.pdf England, N.

Chapman, R., Norton, R. & Paterson, C. (2001) A descriptive outcome study of 291 acupuncture patients. *The European Journal of Oriental Medicine*, 48-53.

Clark-Carter, D. (1997) *Doing Quantitative Psychological Research: From design to report.* East Sussex, UK: Psychology Press.

Clarke, A., Friede, T., Putz, R., Ashdown, J., Martin, S., Blake, A., Adi, Y., Parkinson, J., Flynn, P., Platt, S. & Stewart-Brown, S. (2011) Warwick-Edinburgh Mental Well-being Scale (WEMWBS): validated for teenage school students in England and Scotland. A mixed methods assessment. *BMC Public Health*, 11

Cleeland, C. S. (1989) Brief Pain Scale: Measurement of pain by subjective report. In: C.R. Chapman, J. D. L. E., ed. *Issues in Pain Management*. New York, NY: Raven Press: 391–403.

Cleland, J. A., Price, D. B., Lee, A. J., Gerard, S. & Sharma, A. (2006) A pragmatic, three-arm randomised controlled trial of spiritual healing for asthma in primary care. *British Journal of General Practice*, 56 ((527)): 444-449.

Cohen, J. (1988) Statistical power analysis for the behavioral sciences . Hilsdale. *NJ: Lawrence Earlbaum Associates*, 2

Collins, D. (2003) Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, 12 (3): 229-238.

Cooley, K., Szczurko, O., Perri, D., Mills, E., Bernhardt, B., Zhou, Q. & Seely, D. (2009) Naturopathic care for anxiety: a randomized controlled trial. *PLoS One*, 4 (8)

Costley, C. & Gibbs, P. (2006) Researching others: care as an ethic for practitioner researchers. *Studies in Higher Education*, 31 (1): 88-89.

Couch, A. & Keniston, K. (1960) Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, 60 151-174.

Creswell, J. W. & Clark, V. (2007) *Designing and Conducting Mixed Methods Research*. Thousand Oaks, California, USA: Sage.

Dancey, C. & Reidy, J. (1999) *Statistics Without Maths for Psychology: using SPSS for Windows* ™. London: Prentice Hall.

De Vet, H. C. W., Terwee, C. B., Mokkink, L. B. & Knol, D. L. (2011) *Measurement in Medicine: A practical guide*. Cambridge: Cambridge University Press.

Devlin, N. & Brooks, R. (2017) EQ-5D and the EuroQol Group: Past, Present and Future. *Applied Health Econ Health Policy*, 15 127-137.

Drennan, J. (2002) Cognitive interviewing: verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing*, 42 (1): 57-63.

Edwards, S. (2015) HeartMath: a positive psychology paradigm for promoting psychophysiological and global coherence,. *Journal of Psychology in Africa*, 25 (4): 367-374.

Eisenberg, D., Buring, J., Hrbek, A., Davis, R. & Connelly MT, e. a. (2012) A model of integrative care for low-back pain. . *Journal of Alternative and Complementary Medicine*, *Apr;18* (4): 354-362.

Elden, H., Ostgaard, H. C., Glantz, A., Marciniak, P., Linner, A. C. & Olsen, M. F. (2013) Effects of craniosacral therapy as adjunct to standard treatment for pelvic girdle pain in pregnant women: a multicenter, single blind, randomized controlled trial. *Acta Obstetricia et Gynecologica Scandinavica*, 92 (7): 775-782.

Eton, D., Koffler, K., Cella, D., Eisenstein, A., Astin, J., Pelletier, K. & Riley, D. (2005) Developing a self-report outcome measure for complementary and alternative medicine. *Explore: The Journal of Science and Healing*, 1 (3): 177-185.

Eton, D. T., Temple, L. M. & Koffler, K. (2007) Pilot Validation of a Self-Report Outcome Measure of Complementary and Alternative Medicine. *EXPLORE: The Journal of Science and Healing*, 3 (6): 592-599.

Euroqol Group (1990) EuroQol - a new facilitiy for the measurement of health-related quality of life. *Health Policy*, 16 (3): 199-208.

Euroqol Group (2015) *EQ-5D-5L User Guide*. Available from: https://euroqol.org/eq-5d-instruments/ Rotterdam, The Netherlands:

Fayers, P. M. & Hand, D. J. (1997) Factor analysis, causal indicators and quality of life. *Quality of Life Research*, 6 139-150.

Fayers, P. M. & Machin, D. (2007) *Quality of Life. The assessment, analysis and interpretation of patient-reported outcomes*. Second edition edn. Chichester, England: Wiley.

FDA., U. D. o. H. a. H. S. (2009) Patient-Reported Outcome Measures: Use in medical product development to support labeling claims.

Field, A. (2013) *Discovering Statistics Using IBM SPSS Statistics*. 4th edn. London: SAGE Publications Ltd.

Fitzpatrick, R., Davey, C., Buxton, M. & Jones, D. (1998) *Evaluating patient-based outcome measures for use in clinical trials*. Southampton, UK: Core Research, A.

Gandek, B., Sinclair, S. J., Kosinski, M. & Ware, J. E. (2004) Psychometric Evaluation of the SF-36 ® Health Survey in Medicare Managed Care. *Health Care Financing Review*, *25* (4): 5–25.

Gardner, L. A., Hart, L. K. & Zimmerman, M. B. (2008) Craniosacral still point technique: Exploring its effects in individuals with dementia. *Journal of Gerontological Nursing*, 34 (3): 36-45.

Gilbert, K. (2000) *The emotional nature of qualitative research*. CRC Press.

Glaser, B. G. & Strauss, A. L. (1967) *The Discovery of Grounded Theory: strategies for qualitative research*. New York: Aldine de Gruyter.

Gorecki, C., Brown, J., Cano, S., DL., L., Briggs, M., Coleman, S., Dealey, C., McGinnis, E., Nelson, A., Stubbs, N., Wilson, L. & Nixon, J. (2013) Development and validation of a new patient reported outcome measure for patients with pressure ulcers: the PU-QOL instrument. *Health Quality of Life Outcomes*, 11:95

Grace-Martin, K. (2008-2017) *The Analysis Factor: A short introduction into Factor Analysis Part 1.* [online] Available from: http://www.theanalysisfactor.com (Accessed 1st August 2017).

Group, W. S. (2006) A cross-cultural study of spirituality, religion, and personal beliefs as components of quality of life. *Social Science & Medicine*, 62 (6): 1486-1497.

Guillemin, M. & Gillam, G. (2004) Ethics, Reflexivity, and "Ethically Important Moments" in Research. *Qualitative Inquiry*, 10 (2): 261-280.

Hagins, M. R., A. (2016) Yoga Improves Academic Performance in Urban High School Students Compared to Physical Education: A Randomized Controlled Trial. *Mind, Brain, and Education*, 10 105–116.

Haller, H., Lauche, R., Cramer, H., Rampp, T., Saha, F. J., Ostermann, T. & Dobos, G. (2016) Craniosacral Therapy for the Treatment of Chronic Neck Pain: A Randomized Sham-controlled Trial. *Clinical Journal of Pain*, 32 (5): 441-449.

Hambleton, R. & Jones, R. (1993) Comparison of Classical Test Theory and Item Response theory and their Application to Test Development. *Educational Measurement: Issue and Practice*, (Fall 1993): 38-47.

Health, F. o. P. (2010) *Concepts of Mental and Social Wellbeing*. [online] Available from: http://www.fph.org.uk/concepts_of_mental_and_social_wellbeing (Accessed 12 January).

Healthmeasures (2004) *Patient Reported Outcome information Systems (PROMIS)*. [online] Available from: http://www.healthmeasures.net/explore-measurement-systems/promis (Accessed 27 March).

Henderson, N. (1994) Asking effective focus group questions. *Quirks Marketing Research Review*, 8 (10): 8-35.

Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Bonsel, G. & Badia, X. (2011) Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*, 20 (10): 1727-1736.

Hobart, J. & Cano, S. (2009) *Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods*. London: HMSO, Q. P. a. C. o.

Hoffman, C. J., Ersser, S. J., Hopkinson, J. B. & Nicholls, P. G. (2012) Effectiveness of mindfulness-based stress reduction in mood, breast- and endocrine-related quality of life, and well-being in stage 0 to III breast cancer: a randomized, controlled trial. *Journal of Clinical Oncology,*, *Apr 20* ;30 (12): 1335-1342.

Hungelmann, J., Kenkel-Rossi, E., Klassen, L. & Stollenwerk, R. (1996) Focus on Spiritual Well-Being: Harmonious interconnectedness of mind-body-spirit—Use of the JAREL Spiritual Well-Being Scale: Assessment of spiritual well-being is essential to the health of individuals. *Geriatric Nursing*, 17 (6): 262-266.

Hunt, J., Celho, H., Wider, B., Perry, R., Hung, S., Terry, R. & Ernst, E. (2010.) Complementary and alternative medicine used in England: Results from a national survey. *International Journal of Clinical Practice***.**, 64 1496-1502.

Hunter, J. & Leeder, S. (2013) Patient questionnaires for use in the integrative medicine primary care setting—A systematic literature review. *European Journal of Integrative Medicine*, 5 (3): 194–216.

Huppert, F. & Ruggeri, K. (Forthcoming 2017) Controversies in Well-being: Confronting and resolving the challenges. In: *Oxford Textbook of Public Health*. Oxford, UK: Oxford University Press:

Huppert, F. A. & So, T. T. (2013) Flourishing across Europe: Application of a new conceptual framework for defining well-being. *Social Indicators Research*, 110 (3): 837-861.

Isbell, B. & Carroll, S. (2007) The Effectiveness of Craniosacral Treatment. *TheFulcrum 41, Journal of the Craniosacral Therapy Association, pp 2-5, 2007.The definitive version is available in the 'Articles Online' section at:The Fulcrum*, 41 2-5.

Jackson, A., MacPherson, H. & Hahn, S. (2006) Acupuncture for tinnitus: a series of six n=1 controlled trials. *Complementary Therapies in Medicine*, 14 ((1)): 39-46.

Jackson, D. L., Gillaspy Jr, J. A. & Purc-Stephenson, R. (2009) Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychological methods*, 14 (1): 6.

Kaiser, H. (1960) The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20 141-151.

Keyes, C., Shmotkin, D. & Ryff, C. (2002) Optimizing well-being: The empirical encounter of two traditions. *J Pers Soc Psychol*, 82 1007 - 1022.

Kind, P. & Carr-Hill, R. (1987) The Nottingham Health Profile: a useful tool for epidemiologists? *Social Sciences and Medicine*, 25 905-910.

Kitzinger, J. (1993) Understanding AIDS: researching audience perceptions of Acquired Immune Deficiency Syndrome. In: Eldrigde, J., ed. *Getting the Message: News, Truth, and Power*. London: Taylor & Francis: 271-304.

Kitzinger, J. (1995) Introducing Focus Groups. *British Medical Journal*, 311 299-302.

Korpela, K. M. S., E. Jussila, P. (2016) Nature walks as a part of therapeutic intervention for depression. *Ecopsychology*, 8 (1): 8-15.

Kottner, J., Audige, L., Brorson, S., Donner, A., Gajewski, B., Hrobjartsson, A. (2011) Guidelines for Reporting Reliability and Agreement Studies (GRAAS) were proposed. *Journal of Clinical Epidemiology*, 64 96-106.

Krosnick, J. (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 50 537-567.

Krueger, R. A. & Casey, M. A. (1994) *Focus Groups: A Practical Guide for Applied Research*. 5th Edition, 2015 edn. Sage Publications.

Lee, S., Kim, K. H., Kim, T.H., Kim, J.E., Kim, J.H., Kang, J. W., Kang, K.W., Jung, S.Y., Kim, A.R. & Park, H.J. (2013) Moxibustion for treating knee osteoarthritis: study protocol of a multicentre randomised controlled trial. *BMC complementary and alternative medicine*, 13 (1): 59.

Lewis-Beck, M., Bryman, A. & Futing Liao, T. (2004) *The sage encyclopedia of social science research methods*. Thousand Oaks, CA: Sage Publications, Inc.

Likert, R. A. (1952) A technique for the development of attitude scales. *Educational and psychological measurement*, 12 313-315.

Lohr, K. N. (2002) Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research*, 11 (3): 193-205.

Maheswaran, H., Weich, S., Powell, J. & Stewart-Brown, S. (2012) Evaluating the responsiveness of the Warwick Edinburgh Mental Well-Being Scale (WEMWBS): Group and individual level analysis. *Health and Quality of Life Outcomes*, 10 (1): 156.

Mann, J., Gaylord, S., Faurot, K., Suchindran, C., Coeytaux, R., Wilkinson, L., Coble, R. & Curtis, P. (2012) Craniosacral therapy for migraine: A feasibility study. *BMC Complementary and Alternative Medicine. Conference: International Research Congress on Integrative Medicine and Health*, 12 (no pagination):

Mapi. (2002) *Patient reported outcomes and quality of life database - PROQOLID*. [online] Available from: http://www.proqolid.org/ (Accessed 6 March).

Mataran-Penarrocha, G. A., Castro-Sanchez, A. M., Garcia, G. C., Moreno-Lorenzo, C., Carreno, T. P. & Zafra, M. D. O. (2011) Influence of craniosacral therapy on anxiety, depression and quality of life in patients with fibromyalgia. *Evidence-based Complementary and Alternative Medicine*, 2011 (no pagination) (178769):

McDowell, I. (2006) *Measuring Health. A guide to Rating Scales and Questionnaires*. Third Edition edn. Oxford: Oxford University Press.

McKee, M., Kligler, B., Blank, A., Fletcher, J., Jeffres, A., Casalaina, W., Biryukov, F. (2012) The ADDOPT study (Acupuncture to Decrease Disparities in Outcomes of Pain Treatment): feasibility of offering acupuncture in the community health center setting. *Journal of Alternative and Complementary Medicine*, *Sep;18* . (9): 839-843.

Mendoza, T. R., Wang, X. S. & Cleeland, C. S. e. a. (1999) Brief Fatigue: The rapid assessment of fatigue severity in cancer patients. *Cancer*, 85 1186-1196.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & Group, P. (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS med*, 6 (7): e1000097.

Moinpour, C. M. (1994) Measuring quality of life: an emerging science. *Seminal Oncology*, 21 48–63.

Mokkink, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P. W., Knol, D. K., Bouter, L. & de Vet, H. (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research*, 19 (4): 539-549.

Mokkink, L. e. a. (2010) The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63 (7): 737–745.

Morgan, D. (1988) *Focus Groups as Qualitative Research: Qualitative Research Methods Series 16.* Newbury Park, California: Sage.

MOT. (1992) *Medical Outcomes Trust*. [online] Available from: http://www.outcomestrust. org/instruments.htm (Accessed 6 March).

Mytko, J. & Knight, S. (1999) Body, Mind and Spirit: Towards the integration of Religiosity and Spirituality in cancer quality of life research. *Psycho-Oncology*, 8 439-450.

Nassar-McMillan, S. C. & Borders, L. D. (2002) Use of Focus Groups in Survey Item Development. *The Qualitative Report*, 7 (1): 1-12.

NCHOD. (1998) *PROM Patient reported outcome measurement Group*. [online] Available from: http://phi.uhce.ox.ac.uk/about.php (Accessed 27 March).

NEF (2009) *What is Wellbeing?* [online] Available from: http://www.nationalaccountsofwellbeing.org/learn/what-is-well-being.html (Accessed 5th February).

Ng, S. M., Yau, J. K., Chan, C. L. & Ho, D. Y. F. (2005) The Measurement of Body-Mind-Spirit Well-Being. *Social Work in Health Care*, 41 (1): 33-52.

Nunnally, J. (1967) *Psychometric Theory*. 2 edn. New York: McGraw-Hill.

Nunnally, J. C. & Bernstein, I. H. (1978) Psychometric testing. *New York: McGraw*,

Parsons, M. & Greenwood, J. (2000) - A guide to the use of focus groups in health care research: Part 1. - *Contemporary Nurse*, 9 (- 2): - 169.

Paterson, C. (1996a) Measuring outcomes in primary care: a patient generated measure, MYMOP, compared with the SF-36 health survey. *BMJ*, 312 (7037): 1016-1020.

Paterson, C. (1996b) Measuring outcome in primary care: a patient-generated measure, MYMOP, compared to the SF-36 health survey. *BMJ*, 312 1016 - 1020.

Paterson, C. (2004) Seeking the Patient's Perspective: A Qualitative Assessment of EuroQol, COOP-WONCA Charts and MYMOP. *Quality of Life Research*, 13 (5): 871-881.

Paterson, C., Allen, J. A., Browning, M., Barlow, G. & Ewings, P. (2005) A pilot study of therapeutic massage for people with Parkinson's disease: the added value of user involvement. *Complementary Therapies in Clinical Practice*, 11 (3):

Paterson, C., Britten, N. (2000) In pursuit of patient-centred outcomes: a qualitative evaluation of MYMOP, MeasureYourself Medical Outcome Profile. *Journal of Health Services Research & Policy*, 5 27 - 36.

Paterson, C., Thomas, K., Manasse, A., Cooke, H. & Peace, G. (2007) Measure Yourself Concerns and Wellbeing (MYCaW): An individualised questionnaire for evaluating outcome in cancer support care that includes complementary therapies. *Complementary Therapies in Medicine*, 15 (1): 38-45.

Patton, M. (2015) *Qualitative research and evaluation methods*. 4th edn. Thousand Oaks, California: Sage.

Peterman, A. H., Fitchett, G., Brady, M. J., Hernandez, L. & Cella, D. (2002) Measuring spiritual well-being in people with cancer: The Functional Assessment of Chronic Illness Therapy-Spiritual Well-Being Scale (FACIT-Sp). *Annals of Behavioral Medicine*, 24 49 - 58.

Petrillo, J., Cano, S., McLeod, L. & Coon, C. (2015) Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to evaluate Patient Reported Outcome Measures: A comparison of worked examples. *Value in Health*, 18 25-34.

Polley, M. J., Seers, H. E., Cooke, H. J., Hoffman, C. & Paterson, C. (2007) How to summarise and report written qualitative data from patients: a method for use in cancer support care. *Support Care Cancer*, 15 963-971.

Raviv, G., Shefi, S., Nizani, D. & Achiron, A. (2009) Effect of craniosacral therapy on lower urinary tract signs and symptoms in multiple sclerosis. *Complementary Therapies in Clinical Practice*, 15 (2): 72-75.

Ritenbaugh, C. (2014) *NCCAM Psychometric components of final progress report: developing patient-centered measures for outcomes of CAM therapies*. US:

Ritenbaugh, C., Nichter, M., Kellsy, K. L., Sims, C. M., Bell, I. R., Castaneda, H. M., Elder, C. R., Sutherland, E. G., Verhoef, M. J., Warber, S. L. & Coons, S. J. (2011) Developing a patient-centered outcome measure for complementary and

alternative medicine therapies I: defining content and format. *BMC Complementary and Alternative Medicine*, 11 (135): 17.

Rubin, H. J. & Rubin, S. J. (1995) *Qualitative Interviewing: the art of hearing data*. Thousand Oaks, California: Sage Publications , Inc.

Ryff, C. (1989) Happiness is everything, or is it? Explorations on the meaning of psychological well-being. . *Journal of Personality and Social Psychology.*, 57 (6): 1069-1081.

Seligman, M. (2011) *Flourish: A new understanding of happiness and well-being and how to achieve them.* London, UK: Nicholas Brealey Publishing.

Skillgate, E., Bohman, T., Holm, L. W., Vingård, E. & Alfredsson, L. (2010) The long-term effects of naprapathic manual therapy on back and neck pain-results from a pragmatic randomized controlled trial. *BMC musculoskeletal disorders*, 11 (1): 26.

Speight, J., Barendse, S. & Bradley, C. (2000) The W-BQ 28: further development of the Well-being Questionnaire to include diabetes-specific as well as generic subscales and new stress subscales. *Proceedings of the British Psychological Society*, 8 (1): 21.

Sprangers, M. A. G. & Schwartz, C. E. (1999) Integrating response shift into health-related quality of life research: a theoretical model. *Social Science & Medicine*, 48 (11): 1507-1515.

Stewart, D. W. & Shamdasani, P. N. (1990) *Focus groups: theory and practice*. Sage Publications.

Stewart-Brown, S. (2015a) Measuring wellbeing: What does the Warwick-Edinburgh Mental Well-being Scale have to offer integrated care? *European Journal of Integrative Medicine*, 7 384-388.

Stewart-Brown, S. (2015) “Population level: Wellbeing in the general population”. In: *Wellbeing, Recovery and Mental Health.* Cambridge: Cambridge University Press:

Stewart-Brown, S. (Forthcoming 2017) Resilience and Wellbeing. In: *Oxford Textbook of Public Mental Health*. Oxford, UK: Oxford University Press:

Stewart-Brown, S., Lee, M. S. & Robinson, N. (2015) Issues and Initiatives in outcome measurement for integrative medicine. *European Journal of Integrative Medicine*, 7 319-320.

Streiner, D., Norman, G. & Cairney, J. (2015) *Health Measurement Scales: a practical guide to their development and use*. Fifth Edition edn. Oxford University Press.

Streiner, D. L. & Norman, G. R. (2008) *Health Measurement Scales: a practical guide to their development and use*. Oxford: Oxford University Press.

Sutherland, W. G. (1990) *Teachings in the Science of Osteopathy*. Fort Worth, Texas: Sutherland Cranial Teaching Foundation, Inc.

Tashakkori, A. & Teddlie, C. (2010) Putting the human back in ''human research methodology'': The researcher in mixed methods research. *Journal of Mixed Methods Research*, 4 (4): 271-277.

Tennant, R., Fishwick, R., Platt, S., Joseph, S. & Stewart-Brown, S. (2006) Monitoring positive mental health in Scotland: validating the Affectometer 2 scale and developing the Warwick-Edinburgh Mental Well-being Scale for the UK.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J. & Stewart-Brown, S. (2007) The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health Qual Life Outcomes*, 5

Terwee, C. B., Jansma, E. P., Riphagen, I. I. & de Vet, H. C. (2009) Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*, 18 (8): 1115-1123.

Thomas, K. J., MacPherson, H., Ratcliffe, J., Thorpe, L., Brazier, J., Campbell, M., Fitter, M., Roman, M., Walters, S. & Nicholl, J. P. (2005) Longer term clinical and economic benefits of offering acupuncture care to patients with chronic low back pain. *Health Technology Assessment*, 9 (32, iii-iv, ix-x,): 1-109.

Thompson, E. A., Montgomery, A., Douglas, D. & Reilly, D. (2005) A pilot, randomized, double-blinded, placebo-controlled trial of individualized homeopathy for symptoms of estrogen withdrawal in breast-cancer survivors. *Journal of Alternative and Complementary Medicine,*, 11 (1): 13-20.

Thompson, J. J., Kelly, K. L., Ritenbaugh, C., Hopkins, A. L., Sims, C. M. & Coons, S. J. (2011) Developing a patient-centered outcome measure for complementary and alternative medicine therapies II: Refining content validity through cognitive interviews. *BMC Complementary and Alternative Medicine*, 11 (136): 17.

Thorne, S., Paterson, B., Russell, C. & Schultz, A. (2002) Complementary/alternative medicine in chronic illness as informed self-care decision making. *International Journal of Nursing Studies*, 39 671-683.

van Kampen, D., Willems, W., van Beers, L., Castelein, R., Scholtes, V. & Terwee, C. (2013) Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four shoulder patient-reported outcome measures (PROMs). *Journal of Orthopaedic Surgery and Research*, 8 (40):

Veenhoven, R. (2003) *World Database of Happiness* [online] Available from: http://www1.eur.nl/fsw/happiness/ (Accessed 27 March 2014).

Verhoef, M., Lewith, G., Ritenbaugh, C., Boon, H., Fleishman, S. & Leis, A. (2005a) Complementary and alternative medicine whole systems research: Beyond identification of inadequacies of the RCT. *Complementary Therapies in Medicine*, 13 (3): 206-212.

Verhoef, M., Ware, M., Dryden, T., Gignac, P., Weeks, L. & Kania, A. e. a. (2007) Getting the measures you need: the IN-CAM outcomes database. *Focus on Alternative and Complementary Medicine*, 12 170-171.

Verhoef, M. J., Mulkins, A. & Boon, H. (2005b) Integrative Health Care: How can we determine whether patients benefit? *The Journal of Alternative and Complementary Medicine*, 11 (Supplement 1): S-57-S-65.

Verhoef, M. J., Vanerheyden, L. C., Dryden, T., Mallory, D. & Ware, M. A. (2006) Evaluating complementary and alternative medicine interventions: in search of appropriate patient-centered outcome measures. BMC Complementary and Alternative Medicine 6 (38): Available from: (Accessed

Ware, J. & Sherbourne, C. (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*, 30 473-483.

Ware, J. E., Kosinski, M., Gandek, B., Sundaram, M., Bjorner, J. M., Turner-Bowker, D. M. & Maruish, M. E. (2010) *User's manual for the SF-12v2 Health Survey (2nd ed.)*. Lincoln, RI.:

Ware, J. E., Kosinski, M. & Keller, S. D. (1995) *SF-12: How to score the SF-12 physical and mental health summary scales (2nd ed.).* Boston, MA:

Ware, J. E., Kosinski, M., Turner-Bowker, D. & Gandek, B. (2002) *How to score Version 2 of the SF-12 Health Survey (with a supplement documenting Version 1)*. Lincoln, RI:

Ware, J. E., Sherbourne, C.D. (1992) The MOS 36-item short-form health survey (SF-36). *Medical Care*, 30 473 - 483.

Warren Brown, K. & Ryan, R. M. (2004) Perils and promise in defining and measuring mindfulness: Observations from experience. *Clinical Psychology: Science and Practice*, 11 (3): 242-248.

Watson, D., Clark, L. & Tellegen, A. (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psych*, 6 (54): 1063 - 1070.

Webster, K., Cella, D. & Yost, K. (2003) The Functional Assessment of Chronic Illness Therapy (FACIT) measurement system: properties, applications and interpretations. Health Quality of Life Outcomes, 1 (79):

Weze, C., Leathard, H. L., Grange, J., Tiplady, P., Stevens & G. (2004) Evaluation of healing by gentle touch in 35 clients with cancer. *European Journal of Oncology Nursing*, 8 (1): 40-49.

Wilkin, D., Hallam, L. & Doggett, M.-A. (1992) *Measures of Need and Outcome for Primary Health Care*. United States: Oxford University Press.

Wilson, I. & Cleary, P. (1995) Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *Journal of the American Medical Association*, 273 (1): 59-65.

Winstead-Fry, P. & Schultz, A. (1997) Psychometric assessment of the Functional Assessment of Cancer Therapy-General (FACT-G) scale in a rural sample. *Cancer,*, *79* (12): 2446-2452.

Wong, W., Lam, L. C., Li, R., Ho, S. H., Fai, L. K. & Li, Z. (2011) A comparison of the effectiveness between Western medicine and Chinese medicine outpatient consultations in primary care. *Complementary therapies in medicine*, 19 (5): 264-275.

Yong, A. & Pearce, S. (2013) A Beginner's guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*, 9 (2): 79-94.